

Lukijan suora puhuttelu tietosuojaselosteissa

Suomenkielisten käännösten ja rinnakkaistekstien vertailua

Lotta Tuure
Pro gradu -tutkielma
Kääntämisen ja tulkkauksen maisteriohjelma
Humanistinen tiedekunta
Helsingin yliopisto
Toukokuu 2019

Tiedekunta – Fakultet – Faculty Humanistinen tiedekunta		Koulutusohjelma – Utbildningsprogram – Degree Programme Kääntämisen ja tulkkauksen maisteriohjelma	
Opintosuunta – Studieriktning – Study Track Käännös- ja tulkkausviestintä			
Tekijä – Författare – Author Lotta Tuure			
Työn nimi – Arbetets titel – Title Lukijan suora puhuttelu tietosuojaselosteissa – suomenkielisten käännösten ja rinnakkaistekstien vertailua			
Työn laji – Arbetets art – Level	Aika – Datum – Month and year	Sivumäärä – Sidoantal – Number of pages	
Pro gradu -tutkielma	Toukokuu 2019	46 sivua + englanninkielinen lyhennelmä 7 sivua	
Tiivistelmä – Referat – Abstract			
<p>Tämä pro gradu -tutkielma käsittelee lukijan suoraa puhuttelua suomenkielisissä tietosuojaselosteissa. Tutkimuksen aineistona on kuusi englannista suomeen käännettyä tietosuojaselostetta ja yhdeksän alun perin suomeksi kirjoitettua tietosuojaselostetta eli rinnakkaistekstiä. Tietosuojaselosteella tarkoitetaan tässä tutkimuksessa tekstiä, jossa tietoja keräävä tahon eli rekisterinpitäjän kuva kerättyjen tietojen kohteelle eli rekisteröidylle tämän oikeudet ja sen, mitä tietoja kerätään ja mihin tarkoitukseen. Tietosuojaselosteen sisällöstä säädetään Euroopan unionin yleisessä tietosuojalainsäädännössä, joka tuli voimaan 25.5.2018.</p> <p>Tutkimuksessa tarkastellaan suoraa puhuttelua eli tapoja kohdistaa teksti suoraan lukijalle, ja tutkimuksen tarkoituksena on selvittää, kuinka paljon suoraa puhuttelua käännöksissä ja rinnakkaisteksteissä esiintyy ja millaisia eroja ja yhtäläisyyksiä näiden tekstiryhmien välillä on puhuttelun osalta. Puhuttelukäytännöt ja keinot ilmaista puhuttelua ovat eri kielissä erilaisia, mikä voi heijastua käännettyihin teksteihin interferenssinä eli lähtökielen vaikutuksena kohdekieleen. Aiemmassa tutkimuksessa (kuten Mauranen & Tiittula 2005) on tehty havaintoja esimerkiksi siitä, että englannista suomeen käännettyissä teksteissä erityisesti persoonapronomineja käytetään runsaammin kuin alun perin suomeksi kirjoitetuissa teksteissä, ja tähän perustuu myös tutkimuksen hypoteesi, että aineiston käännöksissä suoraa puhuttelua esiintyy enemmän kuin rinnakkaisteksteissä. Suomen kielessä suoraa puhuttelua voidaan ilmaista persoonapronomineilla <i>sinä</i> ja <i>te</i>, verbinmuodoilla tai possessiivisuffiksilla, minkä lisäksi suomen kielelle on ominaista suorien persoonaviittausten välttäminen. Englannissa suoraan puhutteluun on käytössä ainoastaan yksi pronomini, <i>you</i>, minkä lisäksi puhuttelua voidaan osoittaa imperatiivilla.</p> <p>Tutkimusasetelma on vertaileva, ja aineistoa analysoidaan lähinnä määrällisestä, mutta myös laadullisesta näkökulmasta. Analyysin pohjana käytetään hieman muokattua versiota Eveliina Tolvasen (2016) puhuttelumuotoluokittelusta, jonka perusteella aineistosta poimitut puhuttelumuodot on jaettu neljään luokkaan: (1) toisen persoonan verbinmuodot, (2) <i>sinä</i>-pronomini, (3) possessiivisuffiksit ja (4) imperatiivi. Puhuttelumuotojen poimimisen ja luokittelun jälkeen laskettiin puhuttelumuotojen määrät ja prosenttiosuudet sekä käännös- ja rinnakkaistekstiaineistoissa että yksittäisissä teksteissä. Kummastakin aineistosta laskettiin myös kolme yleisintä sanaa, johon kunkin puhuttelumuotoluokan edustajat liittyvät.</p> <p>Analyysin perusteella voidaan sanoa, että suora puhuttelu oli käännöksissä huomattavasti yleisempää kuin rinnakkaisteksteissä. Käännöksissä puhuttelua oli keskimäärin 12 esiintymää sataa sanaa kohti ja rinnakkaisteksteissä kolme. Puhuttelun määrän lisäksi käännösten ja rinnakkaistekstien välillä oli jonkin verran eroa myös siinä, mitä muotoja käytettiin.</p> <p>Tutkimus antaa osviittaa siitä, miten lukijaa puhutellaan käännöksissä verrattuna rinnakkaisteksteihin, mutta yleistysten tekemiseksi jatkotutkimuksessa tulisi tarkastella laajempaa, muitakin tekstilajeja kattavaa aineistoa. Tutkimustulokset kuitenkin tukevat aiempaa tutkimusta siinä, että persoonaviittauksia esiintyy käännöksissä rinnakkaistekstejä enemmän. Jatkotutkimuksessa voisi tutkia esimerkiksi puhuttelumuotojen kollokaatteja eli sitä, mitkä sanat esiintyvät useimmiten puhuttelumuotojen läheisyydessä.</p>			
Avainsanat – Nyckelord – Keywords, suora puhuttelu, tietosuojaseloste, interferenssi, käännökset, rinnakkaistekstit			
Säilytyspaikka – Förvaringställe – Where deposited Helsingin yliopiston kirjasto			
Muita tietoja – Övriga uppgifter – Additional information			

SISÄLLYS

1 Johdanto	1
2 Puhuttelu	4
2.1 Puhuttelun funktioita	4
2.2 Puhuttelukäytännöt suomen kielessä.....	5
2.3 Puhuttelukäytännöt englannin kielessä	8
2.4 Kohteliaisuus	9
3 Käännetyin suomen ominaispiirteitä	11
3.1 Uniikkiaineet.....	11
3.2 Interferenssi	12
4 Tutkimusaineisto ja -menetelmä	14
4.1 Tietosuojaseloste tekstilajina.....	14
4.1.1 Henkilötietojen kerääminen ja käsittely	15
4.1.2 Rekisteröidyn oikeudet.....	15
4.1.3 Kieleen ja tyyliin liittyvät vaatimukset	15
4.2 Tutkimusaineisto ja sen kerääminen	16
4.3 Tutkimusmenetelmä	18
5 Analyysi ja tulokset	20
5.1 Yleiskatsaus tutkimusaineiston analyysiin.....	20
5.2 Lukijan puhuttelu käännöksissä	27
5.3 Lukijan puhuttelu rinnakkaisteksteissä	32
5.4 Päätelemät	39
6 Lopuksi	42
Lähteet	44

English Summary

LUETTELO TAULUKOISTA

Taulukko 1. Käännös- ja rinnakkaistekstiaineistojen laajuudet	20
Taulukko 2. Käännös- ja rinnakkaistekstiaineistojen laajuus sekä puhuttelumuotoesiintymien kokonaismäärä kummassakin aineistossa.....	21
Taulukko 3. Kunkin puhuttelumuotoluokan esiintymien lukumäärä kussakin aineistossa sekä kunkin puhuttelumuotoluokan prosentuaalinen osuus kyseisen aineiston puhuttelumuotoesiintymien kokonaismäärästä	23
Taulukko 4. Puhuttelumuotoluokkien esiintymät tuhatta sanaa kohti käännöksissä ja rinnakkaisteksteissä	24
Taulukko 5. Eveliina Tolvasen (2016: 13) suomenkielisen Kela-aineiston ja tämän tutkimuksen rinnakkaistekstiaineiston sisältämät puhuttelumuodot laskettuna tuhatta sanaa kohti	26
Taulukko 6. Käännösten laajuus sekä puhuttelumuotoesiintymät kussakin käännöksessä	27
Taulukko 7. Puhuttelumuotoluokkien esiintymät kussakin käännöksessä.....	28
Taulukko 8. Puhuttelumuotoluokkien esiintymät tuhatta sanaa kohti kussakin käännöksessä	28
Taulukko 9. Kolme yleisintä sanaa, johon puhuttelumuoto liittyy, kunkin puhuttelumuotoluokan osalta käännöksissä.	31
Taulukko 10. Rinnakkaistekstien laajuus sekä puhuttelumuotoesiintymien määrä kussakin rinnakkaistekstissä.....	33
Taulukko 11. Puhuttelumuotoluokkien esiintymät kussakin rinnakkaistekstissä.	34
Taulukko 12. Puhuttelumuotoluokkien esiintymät tuhatta sanaa kohti kussakin rinnakkaistekstissä.	34
Taulukko 13. Kolme yleisintä sanaa, johon puhuttelumuoto liittyy, kunkin puhuttelumuotoluokan osalta rinnakkaisteksteissä	37

1 Johdanto

Euroopan unionin yleinen tietosuoja-asetus (EU) 2016/679 eli GDPR (englannin sanoista *General Data Protection Regulation*) tuli voimaan 25.5.2018. Tietosuoja-asetuksesta seurasi muutoksia henkilötietoja keräävien ja käsittelevien tahojen eli *rekisterinpitäjien* velvollisuuksiin informoida henkilötietojen kohteita eli *rekisteröityjä* heidän henkilötietojensa käsittelystä. (Tietosuojavaltuutettu 2019.) Informointivelvoite voidaan täyttää esimerkiksi laatimalla *tietosuojaseloste*, ja tietosuoja-asetus edellyttää muun muassa sitä, että henkilötietojen kohteille toimitettava tieto henkilötietojen käsittelystä on selkeää ja helposti ymmärrettävää tietojen kohteen näkökulmasta arvioituna (mp.). Kielen selkeyteen liittyy Suomessa yleisempiäkin tavoitteita – esimerkiksi hallintolain (434/2003) 9 §:ssä säädetään, että viranomaisen käyttämän kielen tulee olla asiallista, selkeää ja ymmärrettävää. Tämän vuoksi tietosuojaselosteita ja niiden kieltä on myös tärkeää tutkia.

Tämän pro gradu -tutkielman tarkastelun kohteena ovat suomenkieliset tietosuojaselosteet, ja tutkimuksen aineiston muodostavat englannista suomeen käännetyt tietosuojaselosteet sekä alun perin suomeksi kirjoitetut tietosuojaselosteet eli *rinnakkaistekstit*. Rinnakkaisteksteillä tarkoitetaan käännöstieteessä tekstejä, jotka edustavat samaa tekstilajia kuin käännökset mutta jotka on kirjoitettu alun perin kohdekielellä (esimerkiksi Pym 2011: 78). Rinnakkaistekstejä hyödynnetään kääntämisessä apuvälineenä, kun halutaan tietää, millaisia esimerkiksi jonkin tietyn tekstilajin edustajat ovat tietyllä kielellä (mp.). Tässä tutkimuksessa rinnakkaistekstejä hyödynnetään vertailukohtana käännöksille, kun tarkastellaan, millaista lukijan puhuttelu on käännöksissä ja miten se eroaa alun perin suomeksi kirjoitettujen tekstien puhuttelusta.

Tietosuojaselosteella tarkoitetaan tässä tutkimuksessa kaikkia sellaisia tekstejä, joissa henkilötietojen kohteelle kuvataan EU:n yleisen tietosuoja-asetuksen mukaisia tietoja henkilötietojen keräämisestä ja käsittelystä sekä henkilötietojen kohteen oikeuksista. Tällaisten tekstien otsikko voi olla muutakin kuin tietosuojaseloste – esimerkiksi *tietosuojakäytäntö*, *tietosuojaperiaate* tai *tietosuojalauseke* – mutta tässä tutkimuksessa kaikkiin tällaisiin teksteihin viitataan epäselvyyksien välttämiseksi yleisesti termillä tietosuojaseloste. Tietosuojaselosteet ovat yleensä vapaasti saatavilla rekisterinpitäjien verkkosivuilla, ja ne löytyvät useimmiten verkkosivun alalaidasta. Kaikki tämän tutkimuksen aineisto on kerätty nimenomaan yhtiöiden verkkosivuilta, ja täten kaikki analysoidut tekstit ovat siis myös verkkotekstejä.

Tutkimuksen erityinen mielenkiinnon kohde tietosuojaselosteiden osalta on niissä esiintyvä lukijan suora puhuttelu, sillä se on olennainen osa tietosuojaselosteiden viestiä – puhuttelulla osoitetaan, kenelle puhutaan (VISK, määritelmät, s.v. *puhuttelu*) ja täten sillä myös esimerkiksi ilmaistaan, keitä tietosuojaselosteissa mainitut toimenpiteet ja oikeudet koskevat. Tutkimusasetelma on vertaileva: tutkimuksessa vertaillaan lukijan suoran puhuttelun tapoja englannista suomeen käännettyissä tietosuojaselosteissa ja suomeksi alun perin kirjoitetuissa tietosuojaselosteissa eli rinnakkaisteksteissä. Tutkimuksen tavoite on selvittää tietosuojaselosteissa esiintyvää suoraa puhuttelua hakemalla aineiston 15 tekstistä vastausta kysymykseen millä tavoin, jos millään, lukijan suora puhuttelu eroaa englannista suomeen käännettyissä tietosuojaselosteissa verrattuna alun perin suomeksi kirjoitettuihin tietosuojaselosteisiin.

Hypoteesini on, että englanninkielisten lähtötekstien vaikutuksen voi havaita suomenkielisissä käännöksissä esimerkiksi siinä, että suoraa puhuttelua on käännöksissä rinnakkaistekstejä enemmän. Hypoteesini perustuu siihen, että englannista suomeen käännettyissä teksteissä on todettu esiintyvän enemmän persoonaviittauksia, etenkin ensimmäisen ja toisen persoonan pronomineja, kuin alun perin suomeksi kirjoitetuissa teksteissä (Mauranen & Tiittula 2005: 35–36; Tirkkonen-Condit 2002: 213). Koska tutkimuksessa tarkastellaan ainoastaan käännöksiä ja rinnakkaistekstejä, eikä lähtötekstejä, suoria päätelmiä lähtökielen vaikutuksesta käännökseen ei kuitenkaan voida tehdä. Ensisijainen tarkastelun ja mielenkiinnon kohde tässä tutkimuksessa on siis käännetyn suomen ja alkuperäissuomen eli alun perin suomeksi kirjoitetun kielen (Tirkkonen-Condit 2005: 123–124) eroavaisuus puhuttelukäytäntöjen osalta. Tutkimuksessa käsitellään nimenomaan suoraa puhuttelua eli sanamuotoja, joilla teksti kohdistetaan suoraan lukijalle. Tarkastelun kohteen rajausta sai alkunsa tosielämästä, toisin sanoen käännettyjen tietosuojaselosteiden runsaasta sinuttelusta tekemistäni havainnoista, minkä lisäksi jo pro gradu -tutkielman laajuudenkin vuoksi oli mielekästä rajata ulkopuolelle epäsuora lukijaan viittaaminen esimerkiksi sanoilla *asiakas* tai *käyttäjä*.

Johanna Isosävin ja Hanna Lappalaisen (2015: 14–16) mukaan puhuttelua on tutkittu niin aidoissa, nauhoitetuissa vuorovaikutustilanteissa kuin esimerkiksi elokuvissa ja kirjallisuudessa. Englannin kielen puhuttelukäytäntöjä ja niiden kehitystä on historiallisesti tutkittu kirjallisten tekstien osalta paljon muun muassa kaunokirjallisuuden ja kirjeiden näkökulmasta (Jucker & Taavitsainen 2013: 73). Puhuttelun tarkastelu tietosuojaselosteen tyyppisissä virallisemmissä teksteissä näyttäisi sen sijaan jääneen vähemmälle huomiolle. Suomessa Eveliina Tolvanen (2016) on tosin tarkastellut puhuttelua viranomaisteksteissä, jotka

vertautuvat jossakin määrin tietosuojaselosteeseen: molemmissa jokin ei-inhimillinen taho puhuttelee yksittäistä henkilöä standardimuotoisella tekstillä. Niinpä Tolvasen tutkimusta ja tutkimustuloksia voidaan jossakin määrin käyttää vertailukohtana tässä tutkimuksessa.

Tutkielman teoreettisessa taustoituksessa paneudutaan sekä suomen että englannin puhuttelukäytäntöihin ja tapoihin ilmaista puhuttelua. Lisäksi tarkastellaan puhuttelun funktioita ja kohteliaisuuden käsitettä. Tutkimuksessa hyödynnetään myös käännösten kielen ominaispiirteitä käsittelevää kirjallisuutta. Teoreettinen viitekehys esitellään seuraavaksi luvuissa 2 ja 3, joita seuraavat tutkimusaineiston, aineiston genren ja tutkimusmenetelmän esittely luvussa 4, analyysin ja tulosten esittely luvussa 5 sekä kokoava tarkastelu tutkimuksesta luvussa 6.

2 Puhuttelu

Ison suomen kieliopin mukaan puhuttelu on ”sen osoittamista, kenelle sanottu on suunnattu” (VISK, määritelmät, s.v. *puhuttelu*). Tässä pro gradu -tutkielmassa sovelletaan samaa määritelmää. Tämän tutkimuksen tapauksessa *sanotulla* tarkoitetaan kuitenkin *kirjoitettua*, sillä tutkimuksen aineisto koostuu kokonaisuudessaan kirjallisesta materiaalista. Tässä luvussa keskitytään tarkastelemaan puhuttelun funktioita (alaluku 2.1), suomen kielen puhuttelukäytäntöjä (alaluku 2.2), englannin kielen puhuttelukäytäntöjä (alaluku 2.3) sekä lyhyesti myös kohteliaisuutta (alaluku 2.4). *Puhuttelu*-käsitteen ymmärtämiseksi on olennaista tarkastella puhuttelun funktioita ja kohteliaisuutta, koska puhuttelu ei ole ainoastaan sanavalinta. Puhuttelumuodon valintaan liittyy monia tekijöitä, ja valitulla puhuttelumuodolla voi olla monia erilaisia vaikutuksia. Englannin kielen puhuttelukäytäntöjä taas tarkastellaan, koska oletan lähtökielellä olleen jonkin verran vaikutusta tarkasteltuihin käännöksiin.

2.1 Puhuttelun funktioita

Puhuttelulla voidaan nähdä olevan lukuisia erilaisia funktioita. Andreas H. Juckerin ja Irma Taavitsaisen (2013: 73) mukaan puhuttelu on yksi kielen tärkeimmistä interaktiivisista piirteistä, ja Michael Clynen, Catrin Norrbyn ja Jane Warrenin (2009: 1) mukaan puhuttelulla on olennainen rooli siinä, miten ihmisten väliset sosiaaliset suhteet rakentuvat, minkä lisäksi puhuttelutavat voivat heidän mukaansa kertoa yhteiskunnan sosiaalisista rakenteista ja muutoksista. Tolvasen (2016: 8) mukaan puhuttelua taas voidaan tarkastella niin sosiaalisia suhteita heijastavana kuin niitä rakentavanakin ilmiönä. Toisin sanoen siis sen lisäksi, että käytetyistä puhuttelumuodoista voidaan päätellä jotakin vuorovaikutuksen osapuolten välisistä suhteista, puhuttelulla voidaan myös vaikuttaa siihen, millaisiksi nämä suhteet muodostuvat. Esimerkiksi tietosuojaselosteissa esiintyvän lukijan puhuttelun voidaan siis katsoa yhtäältä rakentavan tietoja keräävän tahon (esimerkiksi yhtiön) ja tietojen kohteen (esimerkiksi asiakkaan) välistä suhdetta ja toisaalta heijastavan näiden välillä jo mahdollisesti olemassa olevaa suhdetta. Tietosuojaseloste on kuitenkin siinä mielessä hieman erikoinen tarkastelun kohde puhuttelun osalta, että siinä puhuttelu on yksisuuntaista, eikä osapuolten välille muodostu dialogia.

Myös Friederike Braunin (1988) mukaan erilaisilla puhutteluilmauksilla voidaan rakentaa ja ylläpitää ihmisten välisiä suhteita, minkä lisäksi puhuttelulla voidaan hänen mukaansa ilmaista esimerkiksi etäisyyttä, läheisyyttä tai valtasuhteita. Tämä näkökulma tulee tietosuojaselosteissa

esille, oli se sitten tarkkaan harkittua tai tahatonta. Kun lukijaa puhutellaan suoraan esimerkiksi pronominilla *sinä*, teksti tulee väistämättä hyvin lähelle, sillä sinuttelu yhdistetään suomen kielessä tuttavallisuuteen (Isosävi & Lappalainen 2015: 11). Jos taas lukijaan viitataan jollakin yleisellä puhuttelusanalla, kuten *asiakas* tai *käyttäjä*, teksti ja puhuttelija jäävät lukijalle huomattavasti etäisemmäksi.

Tästä aiheesta kirjoittivat jo vuonna 1960 myös Roger Brown ja Albert Gilman artikkelissaan ”The Pronouns of Power and Solidarity”. Brownin ja Gilmanin mukaan puhuttelu voi olla joko solidaarisuutta ilmentävää tai valtasuhteita ilmentävää. Solidaarisuudesta on kyse silloin, kun puhujat puhuttelevat toisiaan samalla pronominilla eli joko sinuttelevat tai teitittelevät kumpikin. Tällöin puhuttelun sanotaan olevan symmetristä. Valtasuhteita puhuttelu taas ilmentää silloin, kun puhuttelijat puhuttelevat toisiaan eri pronomineilla, eli toinen teitittelee ja toinen sinuttelee. Tällöin puhuttelu on epäsymmetristä. (Brown & Gilman 1960: 255–260.) Puhuttelun symmetrisyyttä tai epäsymmetrisyyttä on mahdotonta arvioida tietosuojaselosteen kaltaisessa tekstissä, sillä kuten todettua, puhuttelu on tällaisissa teksteissä yksisuuntaista. Tietosuojaselosteiden osalta voi kuitenkin tarkastella sitä, mitä muotoja tekstin laatinut taho on päättänyt käyttää – jos tekstissä sinutellaan, tarkoittaako se sitä, että tekstin laatija pyrkii luomaan itsensä ja lukijan välille tuttavallisen suhteen, vai sitä, että tekstin laatija asettaa itsensä lukijan yläpuolelle?

Etäisyyden, läheisyyden, solidaarisuuden ja valtasuhteiden lisäksi sinutteluun ja suoraan puhutteluun ylipäänsä liittyy myös selkeyden näkökulma. Esimerkiksi kun Kela vuonna 2011 päätti siirtyä sinutteluun suuressa osassa viestintäänsä, päätöstä edeltäneissä selvityksissä havaittiin, että sinuttelu lisäsi tekstin selkeyttä huomattavasti (Oikeusasiamies 2014). Vaikka tietosuojaseloste ei olekaan viranomaisen laatima teksti, teksti perustuu viranomaisen antamiin ohjeisiin ja siinä kerrotaan lukuisista lukijaan vaikuttavista asioista samaan tyyliin kuin esimerkiksi juuri Kelan kirjeissä. Tietosuojaselosteessa kerrotaan lukijan oikeuksista ja velvollisuuksista, joten on hyvin tärkeää, että lukija ymmärtää tekstin koskevan juuri häntä. Sinuttelu voi siis olla esimerkiksi valtasuhteiden ilmaisemisen sijaan yksi keino, jolla tekstin selkeydestä varmistutaan.

2.2 Puhuttelukäytännöt suomen kielessä

Puhuttelua ilmaistaan eri kielissä erilaisilla keinoilla (Isosävi & Lappalainen 2015: 11), mikä voi aiheuttaa pulmia kääntäjälle. Suomen kielessä suoraa puhuttelua voidaan ilmaista pelkällä

persoonapronominilla (*sinä* tai *te*), persoonapronominin omistusmuodolla (*sinun* tai *teidän*) verbin persoonapäätteellä (esimerkiksi *käytät*) tai muulla verbinmuodolla (esimerkiksi imperatiivilla, kuten *lue*) tai possessiivisuffiksilla (esimerkiksi *henkilötietosi* tai *rekisteröityessäsi*). Ero tuttavallisen ja muodollisen puhuttelun välille tehdään valitsemalla pronomineista *sinä* ja *te*. (Mp.) Suomen kielessä myös epäsuorat persoonaviittaukset muun muassa kolmannessa persoonassa (esimerkiksi ”*käyttäjällä* on oikeus tarkistaa *häntä* koskevat tiedot”) ovat yleisiä (Hakulinen 2001 [1987]: 209). [Esimerkit poimittu tutkimusaineistosta.]

Lappalaisen (2015: 72) mukaan nyky-Suomessa sinuttelu on jo valtavirtaa. Itse asiassa juuri sinuttelu onkin ollut suomen kielessä alun perin ensisijainen puhuttelun muoto, kunnes ruotsin kielen vaikutuksesta 1700-luvulla alettiin laajemmalti teititteleään ylemmässä asemassa olevia. Teitittelyn rinnalle yleistyi myös kolmannen persoonan ja arvonimien käyttö kunnioitusta osoittavana puhuttelumuotona, mutta 1800-luvun lopulla ja 1900-luvun ensimmäisellä puoliskolla etenkin kolmannen persoonan puhuttelua ja arvonimien käyttöä alettiin vastustaa. 1960- ja 1970-lukujen paikkeilla sinuttelu yleistyi uudelleen, ja myöhemmin sinuttelua on alettu käyttää myös kirjallisissa ja virallisemmissa yhteyksissä, kuten Kelan kirjeissä. Teitittely ei ole kuitenkaan kadonnut suomen kielestä kokonaan, vaan sitä käytetään edelleen vaihtelevissa määrin sinuttelun rinnalla. (Mas. 73–76.) Lappalainen (mas. 80–81) myös referoi Suomessa eri vuosikymmeninä toteutettuja tutkimuksia puhuttelukäytännöistä ja puhutteluun liittyvistä asenteista ja toteaa, että tutkimustenkin perusteella sinuttelu on yleistynyt ajan saatossa ja että siihen suhtaudutaan lähinnä myönteisesti.

Myös Matti Larjavaara (1999) käsittelee *Kielikellossa* julkaistussa artikkelissaan puhuttelun historiaa. Larjavaaran (mp.) mukaan vanhassa puhuttelukulttuurissa teitittelyllä oli kaksi pääasiallista funktiota: puhuteltavan korkeamman statuksen ilmaiseminen ja muun sosiaalisen etäisyyden ilmaiseminen. Sinuttelulla hän taas näkee olleen kolme mahdollista funktiota: sosiaalisen läheisyyden ilmaiseminen, epäkohteliaisuuden ilmaiseminen ja sosiaalisen ylemmyyden ilmaiseminen. Nykypuhuttelussa näitä funktioita ei Larjavaaran (mp.) mukaan enää samassa määrin ole. Teitittelyä käytetään hänen mukaansa nykyään harvemmin, mutta kuitenkin edelleen tietynlaisen sosiaalisen etäisyyden ja kunnioituksen ilmaisuna, kun taas sinuttelu on yleistynyt ja toimii nykyään pitkälti sosiaalisen läheisyyden ilmaisijana. Koska Larjavaaran artikkeli on kuitenkin noin 20 vuotta vanha, sitä ei voi pitää tarkkana kuvauksena 2010-luvun lopun nykykielestä. Hänen kuvauksensa antavat kuitenkin suuntaviivoja siitä, miten kieli on puhuttelun osalta ajan mittaan kehittynyt. En ole myöskään löytänyt uudemmassa

aihetta käsittelevästä kirjallisuudesta väitteitä, jotka todistaisivat Larjavaaran havainnot epäsopiviksi nykypäivään.

Larjavaara (1999) vie yleistyksensä kuitenkin varsin pitkälle väittäessään, että lähityyppinen kohteliaisuus ”toimii aina ja kaikkialla”. Lähityyppisellä hän tarkoittaa sellaista kohteliaisuutta (ja puhuttelua), jossa kommunikaation osapuolet puhuttelevat toisiaan kuin kavereitaan – vapaamuotoisesti ja tuttavallisesti. Larjavaara toteaa itsekin, että tällainen kommunikaatio on tyyppillistä etenkin Amerikassa, mutta nyt myös Suomessa. Tällaista kehitystä on toki havaittavissa Suomessa ainakin jossakin määrin (ks. esimerkiksi Lappalainen 2015: 72 tasavallan presidentin puhuttelusta televisiohaastattelussa), eikä amerikkalaisen kulttuurin ja englannin kielen vaikutusta suomalaiseen kulttuuriin ja suomen kieleen voi kiistää. En kuitenkaan uskaltaisi väittää, että Suomessa ja suomen kielessä lähikohteliaisuus toimii tilanteessa kuin tilanteessa, mistä onkin todisteena muun muassa Kelan sinuttelupäätöksestä tehty kantelu (Oikeusasiamies 2014). Tutkielmani kannalta on joka tapauksessa mielenkiintoista pohtia, heijastuuko Larjavaaran kuvailema muutos yhteiskunnassa, puhuttelunormeissa ja asenteissa myös tutkimusaineistooni.

Kuten todettua, persoonapronominien *sinä* ja *te* käyttö ei kuitenkaan ole suomen kielessä puhuttelun osoittamiseksi välttämätöntä, vaan puhuttelua voidaan ilmaista myös esimerkiksi pelkällä verbinmuodolla tai substantiiviin liittyvällä possessiivisuffiksilla (Isosävi & Lappalainen 2015: 12). Lappalaisen (2015: 82) mukaan pronominiton puhuttelumuoto on persoonapronominilla sinuttelua lievempi muoto, ja se voidaan nähdä teitittelyn ja persoonapronominilla sinuttelun välimuotona tai jonkinlaisena kompromissina. Myöhemmin luvussa 5 tarkastelen eri puhuttelumuotojen esiintymistä tutkimusaineistossani.

Pronominien sekä verbien ja substantiivien taivutusmuotojen lisäksi puhuttelua voidaan ilmaista epäsuorasti lukemattomilla kolmannen persoonan puhuttelusanoilla, joita voivat olla esimerkiksi tittelit, sukulaisuussanat tai vaikkapa haukkumasanat. Tällaisten puhuttelusanojen käyttö ei kuitenkaan ole koskaan ollut suomessa yleistä, vaan nykyään niitä käytetään lähinnä esimerkiksi humoristisessa mielessä. (Isosävi & Lappalainen 2015: 12; Lappalainen 2015: 84–85.) Epäsuoraa puhuttelua tarkastellaan tutkimusaineistoni osalta lyhyesti analyysin yhteydessä alaluvussa 5.3.

Suomen kielessä yleistä on myös puhuttelun kiertäminen kokonaan esimerkiksi nollapersoonaa käyttämällä. Muun muassa Auli Hakulisen (2001 [1987]: 209) mukaan eksplisiittisten

persoonaviittausten välttäminen on ”syvästi juurtunut [suomen] kielen rakenteisiin”. Hakulisen (mas. 211–216; esimerkit Hakulisen, muotoilu Tuuren) puhuttelu voidaan kiertää esimerkiksi käyttämällä passiivia (esimerkiksi ”Sitä *olla*n niin salaperäisen näköisiä!”) tai geneeristä kolmatta persoonaa (esimerkiksi ”Jos noin *tekee*, väsy nopeasti.”). Lisäksi useissa puhuttelutilanteissa suoria ja epäsuoria puhuttelutyylejä käytetään sekaisin, samoin kuin joskus sinuttelua ja teitittelyä (Lappalainen 2015: 82–83, 91–100). Tähän liittyy myös Lappalaisen (mas. 100) puhuttelututkimuksen tärkein tulos – hänen mukaansa ”Suomessa ei ole selviä puhuttelunormeja, minkä vuoksi moni suomalainen kokee sopivan muodon valinnan vaikeaksi”. Vaikka Lappalaisen tutkimus keskittyykin pitkälti sinuttelun ja teitittelyn vaihteluun asiakaspalvelutilanteiden keskusteluissa, hänen johtopäätöksensä on myös tämän tutkimuksen kannalta mielenkiintoinen. Kun suomen kielessä ei ole vakiintunutta puhuttelunormistoa, kääntäjälläkään ei ole valmista vastausta siihen, miten esimerkiksi englannin kielen *you* tulisi suomeksi kääntää.

2.3 Puhuttelukäytännöt englannin kielessä

Englannin kielessä puhuttelua voidaan ilmaista pääasiassa kahdella tavalla – pronominilla (*you*) tai substantiivilla (esimerkiksi *miss*). Usein käytetään myös näiden kahden yhdistelmää (esimerkiksi *you fool*). (Jucker & Taavitsainen 2013: 73.; esimerkit Tuure) Lisäksi suomen tavoin englannissakin voidaan osoittaa puhuttelu imperatiivilla (Jespersen 1952: 468).

Pronominipuhuttelua voidaan englannissa nykyään ilmaista käytännössä ainoastaan pronominilla *you* (Isosävi & Lappalainen 2015: 12; Clyne, Norrby & Warren 2009: 4). Englannissakin on kuitenkin aikaisemmin ollut puhutteluun käytettävissä kaksi persoonapronominia: *ye* ja *thou*. Karkeasti yleistäen *ye*-pronominia käytettiin muodollisissa tilanteissa ja *thou*-pronominia tuttavallisemmissa tilanteissa, mutta puhuttelunormit olivat vanhassa englannissa monimutkaiset, eivätkä ne vastanneet täydellisesti minkään muun kielen puhuttelunormeja. Alun perin pronomini *thou* oli yleisin puhuttelun muoto, mutta jo 1600-luvulla sitä käytettiin loukkaavana ja alentavana ilmauksena. 1700-luvulle tultaessa kahden pronominin järjestelmä oli käytännössä hävinnyt, ja käytössä oli enää pronominista *ye* hioutunut muoto *you*. (Jucker & Taavitsainen 2013: 76–83.) Suomella ja englannilla on siis puhuttelun osalta myös jotakin yhteistä – molemmissa kielissä puhuttelutavat ja -normit ovat kehittyneet ja muuttuneet ajan saatossa runsaasti.

Nykyenglannissa puhutteluilmausten ja -sanojen käyttöä suositaan enemmän kuin suomessa (Isosävi & Lappalainen 2015: 11). *You*-pronominia käytetään paljon, ja se on käytännössä täysin neutraali ilmaus. Virallisen tai tuttavallisen sävyn saavuttamiseksi taas käytetään muita keinoja, kuten titteleitä tai etunimeä. (Clyne, Norrby & Warren 2009: 4, 39.) Tässä kohtaa englanti siis eroaa suomesta varsin merkittävästi – suomessa ero muodollisen ja tuttavallisen välille voidaan tehdä pronominivalinnalla, eikä titteleitä juurikaan käytetä, kun taas englannissa pronomini on aina sama ja neutraali, ja sävyyn vaikutetaan muilla valinnoilla. Englannin kieli eroaa suomen kielestä puhuttelukäytännöiltään siis paljon esimerkiksi juuri tittelien käytössä (Nevala 2015: 280–284). Englannin kielessä tittelin ja sukunimen yhdistelmää käytetään yhä paljon, ja esimerkiksi opiskelijat tai oppilaat puhuttelevat usein opettajaansa tällä yhdistelmällä. Ainoastaan silloin, jos opettaja on hyvin tuttu, saatetaan käyttää etunimeä. Tuttavallisemmissa keskusteluissa etunimiä taas voidaan käyttää runsaastikin. (Mp.)

2.4 Kohteliaisuus

Puhuttelu liittyy vahvasti kohteliaisuusnormeihin, joten puhuttelun yhteydessä otetaan usein esille myös kohteliaisuuden käsite (Isosävi & Lappalainen 2015: 9–22). Kohteliaisuustutkimuksessa on ollut suuri rooli muun muassa Penelope Brownilla ja Stephen C. Levinsonilla, joiden kohteliaisuusteoria perustuu *kasvojen* käsitteelle. Brownin ja Levinsonin (1987) teoriassa puhujalla on kahdet kasvot – positiiviset ja negatiiviset. Positiivisiin kasvoihin kuuluu henkilön omakuva tai ”persoonallisuus”, jolle halutaan kanssaihminen hyväksyntä, kun taas negatiivisiin kasvoihin kuuluvat etäisyys ja oman toiminnan vapaus. Brownin ja Levinsonin mukaan kohteliaisuus perustuu siihen, että ihmisillä on tarve kunnioittaa omia ja toisten kasvoja kanssakäymisessään. (Mts. 61.)

Positiivisiin kasvoihin liittyy *positiivinen kohteliaisuus*, joka korostaa puhujien samanlaisuutta ja tuttuutta. Negatiivisiin kasvoihin taas liittyy *negatiivinen kohteliaisuus*, joka ylläpitää puhujien välistä etäisyyttä ja välimatkaa. Erilaisilla puhuttelumuodoilla on erilainen vaikutus positiivisiin ja negatiivisiin kasvoihin. Tämä tarkoittaa sitä, että esimerkiksi liian tuttavallinen puhuttelumuoto voi ikään kuin uhata negatiivisia kasvoja, koska se koetaan loukkaukseksi tai muuten vääräksi. Kohteliaisuudessa onkin kyse nimenomaan puhuteltavan odotuksista, ja kun näistä odotuksista poiketaan, kasvot joutuvat uhatuksi. (Brown & Levinson 1987: 65–70.) Kuten puhuttelukäytännöt, myös kohteliaisuusnormit ovat eri kielissä ja maissa erilaiset (mts. 13), mikä on myös kääntämisen kannalta mielenkiintoista ja jokseenkin ongelmallista.

Lienee todennäköistä, että tämän pro gradu -tutkielman tutkimusaineiston tekstejä laadittaessa ei ole pohdittu kohteliaisuusteoriaa ja otettu tietoisesti huomioon lukijan kasvoja. On kuitenkin mahdollista, ellei todennäköistäkin, että alkuperäisiä tekstejä laadittaessa on pohdittu yleisellä tasolla sitä, millainen puhuttelu olisi tekstissä sopivaa ja kohteliasta. Sitä, ovatko tutkimusaineistooni valittujen käännösten laatijat pohtineet kohteliaisuutta, emme tietenkään voi tietää. Mikäli kääntäjä ei kuitenkaan huomioi kohteliaisuusnormeja, vaan kääntää esimerkiksi puhuttelun sellaisenaan suoraan kohdekielelle, hän voi siis tulla uhanneeksi käännöksen lukijan kasvoja.

Larjavaara (1999) vertaa positiivista ja negatiivista kohteliaisuutta lähi- ja etäkohteliaisuuteen. Karkeasti yleistettynä lähikohteliaisuudessa ollaan tuttavallisia ja etäkohteliaisuudessa muodollisempia ja etäisempiä. Englanti, etenkin amerikanenglanti, suosii lähikohteliaisuutta (mp.), mutta suomen kielen luokittelu jompaankumpaan ei ole aivan yhtä selkeää. Englannissa sinutellaan ja puhutellaan enemmän kuin suomessa (ks. alaluvut 2.2 ja 2.3), joten kohteliaisuusnormit ja -odotukset eivät ole täysin samanlaiset. Koska suomen kieli suosii epäsuoraa puhuttelua ja persoonaviittausten välttämistä (Hakulinen 2001 [1987]: 209), sen voisi kenties nähdä näiltä osin suosivan etäkohteliaisuutta. Tällaisessa tapauksessa kohteliaisuusnormit voisivat siis tulla rikotuksi tai vähintään joutua koetukselle, mikäli esimerkiksi englannista suomeen käännettäessä puhuttelumuodot siirtyisivät sellaisenaan suomenkieliseen käännökseen. Kohteliaisuuden käsite on täten siis mielenkiintoinen myös tietosuojaselosteiden ja niiden käännösten näkökulmasta katsottuna.

3 Käännetyn suomen ominaispiirteitä

Tässä luvussa tarkastellaan joitakin tutkimusaineistoni kannalta olennaisia käännösten ominaispiirteitä erityisesti suomen kielen näkökulmasta. Tämän tutkimuksen näkökulmasta näitä piirteitä ovat uniikkiaineokset ja interferenssi, joita käsitellään tarkemmin alaluvuissa 3.1 ja 3.2.

3.1 Uniikkiaineokset

Uniikkiaineoksilla tarkoitetaan sellaisia kielen aineksia, jotka ovat ainutlaatuisia ja tietylle kielelle spesifejä (Tirkkonen-Condit 2005: 123). Tällaisia aineksia on Sonja Tirkkonen-Conditin (2004: 177) mukaan kaikissa kielissä. Käännöksistä puhuttaessa uniikkiaineokset ovat siis sellaisia kohdekielen elementtejä, joiden olemassaolosta lähtökieli ei muistuta. Tirkkonen-Conditin (2005: 123) mukaan suomessa tällaisia elementtejä ovat indoeurooppalaisiin kieliin verrattuna esimerkiksi liitepartikkelit (kuten *-kin*, *-kAAn*, *-pA*). Siispä kun kääntäjä kääntää tekstiä esimerkiksi englannista suomeen, hän ei käytä liitepartikkeleita yhtä paljon kuin hän käyttäisi suoraan suomeksi kirjoittaessaan, sillä lähtötekstissä ei ole vastaavaa elementtiä, eikä se näin ollen tarjoa kääntäjälle kimmoketta hyödyntää kyseistä elementtiä. Tirkkonen-Condit (mas. 124) on myös havainnut, että silloin, kun uniikkiaineoksia esiintyy käännöksissä, niitä saatetaan käyttää eri tavalla kuin rinnakkaisteksteissä.

Tirkkonen-Condit on tutkinut uniikkiaineoksia suuren korpuksen avulla useista eri tekstilajeista, kuten kaunokirjallisuudesta, lastenkirjallisuudesta ja asiateksteistä (Tirkkonen-Condit 2005: 125). Korpustutkimuksessaan *-kin*-partikkelin käytöstä käännöksissä ja alun perin suomeksi kirjoitetuissa teksteissä Tirkkonen-Condit havaitsi, että partikkelin käytössä oli suurta vaihtelua yksittäisten kirjoittajien välillä. Lisäksi hän teki sellaisen huomion, että vaihtelu oli suurempaa alun perin suomeksi kirjoitetuissa teksteissä kuin käännöksissä. (Mp.) Tirkkonen-Conditin tutkimuksessa selvisi myös, että eri tekstilajienkin välillä oli merkittäviä eroja (mas. 127).

Tutkimuksessani uniikkiaineokset ovat relevantteja, sillä suomen kielen puhuttelukeinoihin kuuluu sellaisia elementtejä, joita englannin kielessä ei ole käytettävissä. Edellä alaluvuissa 2.2 ja 2.3 tarkasteltiin suomen ja englannin kielten tapoja ilmaista puhuttelua. Lyhyesti sanottuna suomen kielessä suoraa puhuttelua voidaan siis ilmaista persoonapronominilla, verbin taivutuksella tai possessiivisuffiksilla. Englannin kielessä suoraan puhutteluun on käytössä

käytännössä ainoastaan yksi pronomini, *you*. Verbin persoonapäätteiden ja possessiivisuffiksien voitaisiin siis katsoa olevan englantiin verrattuna suomen uniikkiaineisia.

Uniikkiaines-käsitteen voisi suoran puhuttelun osalta ajatella myös käänteisesti: koska englannin kielessä pronominia *you* käytetään runsaammin kuin suomen kielessä pronomineja *sinä* tai *te*, lähtökieli saattaakin antaa kimmokkeen käyttäjä sinuttelua enemmän kuin suomen kielessä on normaalia. Tätä ilmiötä kutsutaan yleisemmin myös *interferenssiksi*, jota siirrytään käsittelemään seuraavaksi alaluvussa 3.2.

3.2 Interferenssi

Perinteisenä interferenssin määritelmänä pidetään Anna Mauraen (2004: 66) mukaan Uriel Weinreichin (Weinreich 1979 [1953]: 1) alun perin vuonna 1953 julkaistua määritelmää:

[t]hose instances of deviation from the norms of either language which occur in the speech of bilinguals as a result of their familiarity with more than one language, i.e. as a result of language contact.

Interferenssi tarkoittaa siis yksinkertaisesti sanottuna yhden kielen vaikutusta toiseen kieleen. Vaikka Weinreichin määritelmässä puhutaan interferenssistä nimenomaan kaksikielisten henkilöiden kielenkäytön ilmiönä, määritelmää on alettu soveltamaan myös käännöstieteessä. Weinreichin määritelmän jälkeen interferenssistä ja sen suhteesta käännöstieteeseen on esitetty paljonkin eriäviä mielipiteitä. Esimerkiksi Mona Bakerin (ks. esim. Baker 1993) mukaan interferenssi ei lukeudu *käännösuniversaaleihin* eli sellaisiin elementteihin, jotka ovat aina tai lähes aina läsnä käännöksissä kieliparista riippumatta. Toisaalta taas esimerkiksi Gideon Toury (ks. esim. Toury 1995), joka on kirjoittanut käännöksille ominaisista piirteistä jo 1980-luvulla, on esittänyt, että interferenssi olisi yksi kääntämisen ”laeista”. (Mauranen & Kujamäki 2004: 1; Mauranen 2004: 66.)

Tässä pro gradu -tutkielmassa ei ole kuitenkaan tarkoituksenmukaista pohtia sitä, kuuluuko interferenssi käännösuniversaaleihin vai ei. Olennaista on tiedostaa, että ilmiö on olemassa, sillä se on yksi mahdollinen tekijä, joka voi vaikuttaa käännöksissä esiintyvään puhutteluun. Mauraen (2004: 69) mukaan interferenssi voi nimenomaan myös olla kieliparikohtaista, jolloin sitä ei voisi termin varsinaisessa merkityksessä laskea käännösuniversaaliksi. Mauraen (mas. 67) mukaan yleinen käsitys on se, että interferenssiä esiintyy useimmiten syntaksin eli lauseen tasolla, mutta interferenssiä on kuitenkin havaittu kaikilla muillakin tekstin tasoilla. Tutkimusaineistossani interferenssiä voisi olettaa esiintyvän sanan tasolla, koska

tarkasteltavana on yksittäisiä ilmaisuja esimerkiksi lauserakenteiden sijaan. Tämän oletuksen absoluuttinen vahvistaminen edellyttäisi kuitenkin myös lähtötekstien tarkastelua ja vertaamista käännöksiin, joten alkutekstin aiheuttamasta interferenssistä ei tässä tutkimuksessa voida saada varmuutta. Vaikka kysymys onkin mielenkiintoinen, se on rajattu tämän tutkimuksen ulkopuolelle, sillä tutkimuksen tarkoitus on ensisijaisesti selvittää käännösten ja rinnakkaistekstien eroja, ei käännösratkaisujen syitä.

4 Tutkimusaineisto ja -menetelmä

Tutkimuksen aineiston muodostavat englannista suomeen käännettyt tietosuojaselosteet sekä alun perin suomeksi kirjoitetut tietosuojaselosteet eli rinnakkaistekstit. Seuraavissa alaluvuissa perehdytään tietosuojaselosteeseen tekstilajina ja kuvataan tutkimuksen aineisto tarkemmin.

4.1 Tietosuojaseloste tekstilajina

EU:n yleinen tietosuoja-asetus (EU) 2016/679 eli GDPR (englannin sanoista *General Data Protection Regulation*) tuli voimaan 25.5.2018. Tietosuoja-asetuksesta seurasi muutoksia henkilötietoja keräävien ja käsittelevien tahojen eli *rekisterinpitäjien* velvollisuuksiin informoida henkilötietojen kohteita eli *rekisteröityjä* heidän henkilötietojensa käsittelystä. Ennen tietosuoja-asetusta henkilötietojen käsittelyyn sovellettiin Suomessa henkilötietolakiä, jonka mukaan rekisterinpitäjien tuli laatia tietosuojaseloste, joka sisälsi lain vaatimat tiedot. Tietosuoja-asetuksen myötä tietosuojaselosteita ei enää vaadita, vaan rekisteröityjen informoinnin tulee täyttää uudenlaiset vaatimukset. Vaikka tietosuoja-asetus ei vaadi nimenomaisesti tietosuojaselosteen laatimista (2016/679, 12 artikla 1 kohta), se kuitenkin edellyttää, että kaikki informaatio on myös rekisterinpitäjän verkkosivuilla saatavissa. (Tietosuojavaaltuutettu 2019.) Tästä syystä useilla toimijoilla on käytännössä siis edelleen verkkosivuillaan tietosuojaseloste tai jotenkin muuten nimetty (esimerkiksi *tietosuojakäytäntö*) teksti, josta käyvät ilmi tietosuoja-asetuksen vaatimusten mukaiset tiedot siitä, miten ja mihin tarkoitukseen henkilötietoja käsitellään ja mitkä ovat rekisteröidyn oikeudet. Tässä tutkimuksessa kaikkiin tällaisiin teksteihin viitataan yksiselitteisyyden nimissä yleisesti termillä *tietosuojaseloste*.

Tietosuojaseloste on hyvin omanlaisensa teksti, minkä vuoksi sen voisi nähdä muodostavan kokonaan oman tekstilajinsa. Tietosuojaselosteiden sisällön määrittelee pitkälti lainsäädäntö, tarkemmin sanottuna tietosuoja-asetus (2016/679, artikkelit 13–22), joten sisällöiltään tekstit ovat hyvin samansuuntaisia. Myös asiakirjan muotoon ja kieleen liittyen on annettu tiettyjä viranomaisohjeita (Tietosuojavaaltuutettu 2019), minkä voidaan olettaa vaikuttavan tietosuojaselosteiden muotoon yhtenäistävästi. Tietosuojaselosteet ovat myös pääasiallisesti verkkotekstejä, eli tekstilajin voisi katsoa olevan jonkinlainen yhdistelmä oikeusalan tekstiä, viranomaisviestintää ja verkkotekstiä. Seuraavat alaluvut käsittelevät joitakin olennaisimpia tietosuojaselosteen sisältöön, muotoon ja kieleen liittyviä vaatimuksia.

4.1.1 Henkilötietojen kerääminen ja käsittely

Tietosuojaselosteen tarkoituksena on kertoa rekisteröidylle tämän henkilötietojen keräämisestä ja käsittelystä. Jokaisesta tietosuojaselosteesta on siis löydyttävä tieto esimerkiksi siitä, mitä tietoja kerätään ja käsitellään, mihin tarkoitukseen ja millä perusteella, luovutetaanko tietoja kolmansille osapuolille ja kauanko tietoja säilytetään. (Tietosuojavalvottu 2019.)

4.1.2 Rekisteröidyn oikeudet

Jokaisesta tietosuojaselosteesta on löydyttävä myös tieto rekisteröidyn oikeuksista, joita tietuoja-asetuksen mukaan ovat esimerkiksi oikeus saada pääsy omiin tietoihinsa, oikeus vaatia omien tietojensa oikaisemista tai poistamista ja oikeus rajoittaa omien tietojensa käsittelyä. Lisäksi selosteessa on oltava tieto siitä, miten rekisteröity voi käyttää oikeuksiaan. (Tietosuojavalvottu 2019.)

4.1.3 Kieleen ja tyyliin liittyvät vaatimukset

Rekisterinpitäjän on tietuoja-asetuksen 12 artiklan 1 kohdan mukaan pyrittävä tarjoamaan tietosuojaselosteen tiedot johdonmukaisesti, helposti omaksuttavina kokonaisuuksina ja ymmärrettävää, yksinkertaista ja selkeää kieltä käyttäen. Kielen ymmärrettävyyttä ja johdonmukaisuutta tulee arvioida nimenomaan rekisteröidyn näkökulmasta. Tietosuojavalvottun ohjeistuksen mukaan erityisesti henkilötietojen käsittelyn tarkoitus ja sen seuraukset rekisteröidylle tulee kuvata erityisen selkeästi. (Tietosuojavalvottu 2019.)

Kuten aiemmin luvussa 1 mainittiin, sinuttelun on todettu tuovan tekstiin selkeyttä, joten tietosuojaselosteiden sisältämä puhuttelu on tästäkin näkökulmasta hyvin olennaista. Kaikki edellä kuvatut seikat ovat myös käännöksen ja kääntämisen kannalta hyvin tärkeitä, sillä eri kielissä on eri keinot ilmaista asioita. Tietosuojavalvottun ohjeissa muistutetaan toisaalta myös siitä, että eri kieliversioiden tulisi vastata toisiaan, mikä saattaisi taas osaltaan vaikuttaa käännösratkaisuihin esimerkiksi siinä, että lähtötekstin rakenteita noudatetaan hyvin läheisesti.

Edellä tässä luvussa kuvatuista syistä on tärkeää tutkia, eroavatko tietosuojaselosteiden käännökset ja rinnakkaistekstit toisistaan jollakin olennaisella tavalla. Tässä tutkimuksessa asiaa tutkitaan suoran puhuttelun näkökulmasta, sillä tietosuojaselosteessa on tärkeää osoittaa, keitä tekstissä kuvatut toimenpiteet koskevat – suora puhuttelu on tähän yksi keino. Seuraavassa osiossa syvennyttään tarkemmin tutkimuksen aineistoon ja sen keräämiseen.

4.2 Tutkimusaineisto ja sen kerääminen

Tutkimusta varten kerättiin kuusi englannista suomeen käännettyä tietosuojaselostetta ja yhdeksän alun perin suomeksi kirjoitettua tietosuojaselostetta. Kerättyjen selosteiden valintaan vaikuttivat rekisterinpitäjän kotipaikka ja pääasiallinen toimintakieli sekä tietosuojaselosteen pituus ja muoto. Käännösaineistoon valittiin tekstejä yhdysvaltalaisilta yhtiöiltä, joiden pääsääntöinen toimintakieli on englanti, ja rinnakkaistekstiaineistoon tekstejä suomalaisilta yhtiöiltä, joiden pääasiallinen toimintakieli on suomi. Näin pyrittiin varmistamaan se, että käännökset olisivat mahdollisimman suurella todennäköisyydellä käännöksiä englannista ja rinnakkaistekstit alun perin suomeksi kirjoitettuja. Selosteiden valintaan vaikutti myös selosteen muoto ja pituus, sillä vertailukelpoisuuden takaamiseksi aineisto pyrittiin koostamaan mahdollisimman samanmittaisista teksteistä. Tästä huolimatta käännösaineiston tekstit ovat lähes poikkeuksetta rinnakkaistekstejään jonkin verran pidempiä, sillä suurten, kansainvälisten yhtiöiden tietosuojaselosteet ovat esimerkiksi kerättyjen tietotyyppien suuren määrän vuoksi väistämättäkin pidempiä kuin pienempien suomalaisyhtiöiden. Muodon osalta vaatimuksena taas oli, että teksti oli kokonaisia lauseita, eikä esimerkiksi taulukkomuodossa.

Lopullinen, analysoitu tutkimusaineisto ei koostu kokonaisista tietosuojaselosteista, vaan vertailukelpoisuuden takaamiseksi jokaisesta tarkasteltavasta tietosuojaselosteesta valittiin tutkittavaksi tietyt osat. Käännösaineiston muodostavat yhdysvaltalaisen yhtiöiden Applen, Netflixin, Microsoftin, Facebookin, McDonald'sin ja Electronic Artsin tietosuojaselosteista poimitut otteet ja rinnakkaistekstiaineiston muodostavat suomalaisten Atrian, Sanoman, Helsingin seudun liikenteen (HSL), Alkon, Verkkokauppa.comin, Olvin, Evidensian, Mustin & Mirrin sekä K-Plussan tietosuojaselosteista poimitut otteet. Käännettyjen tekstiotteiden yhteispituus on 5 142 sanaa (48 582 merkkiä) ja rinnakkaistekstiotteiden yhteispituus 4 707 sanaa (47 476 merkkiä). Tietosuoja-asetus määrää tietosuojaselosteiden sisällöstä hyvin tarkasti (ks. alaluku 4.1 yllä), mistä johtuen tietyt aiheet toistuvat kaikissa teksteissä. Tietosuojaselosteissa on kuitenkin myös rekisterinpitäjäkohtaisia sisältöjä. Niinpä aineisto on yhdenmukaisuuden ja vertailukelpoisuuden takaamiseksi rajattu sellaisiin osiin, jotka toistuvat pääsääntöisesti kaikissa tietosuojaselosteissa. Osat on valittu sillä perusteella, mitkä asiat ovat tietosuojavaaltuutetun (2019) mukaan tietosuojaselosteiden olennaisimpia osia. Jokaisesta valitusta selosteesta poimittiin tällä perusteella:

- (1) kappaleet, joissa kuvataan, **mitä henkilötietoja kerätään ja käsitellään**
- (2) kappaleet, joissa kuvataan, **mihin tarkoitukseen tiedot kerätään**
- (3) kappaleet, joissa kuvataan **rekisteröidyn oikeudet**, sekä
- (4) mahdollinen **johdantokappale**.

Pääsääntöisesti kaikista tutkimusta varten kerätyistä tietosuojaselosteista löytyivät nämä osat, mutta joissakin tapauksissa joko johdanto tai rekisteröidyn oikeuksia käsittelevä kappale puuttui, ja tällöin kerättiin normaalisti muut edellä luetelluista osista. Jaottelu ei myöskään kaikissa tapauksissa ollut täysin yksiselitteistä esimerkiksi tiedon ryhmittelyn vuoksi, mistä johtuen tarkasteltujen osioiden sisällöt eivät kaikilta osin vastaa toisiaan täydellisesti. Tarkasteltavaksi on kuitenkin pyritty rajaamaan kaikista teksteistä mahdollisimman hyvin toisiaan ja yllä kuvattuja kohtia 1–4 vastaavat osiot. Kuten on jo aiemmin mainittu, tietosuojaselosteiden pituuksissa on myös rekisterinpitäjäkohtaisia eroja, mistä johtuen rinnakkaistekstejä on lukumääräisesti enemmän kuin käännöksiä. Sana- ja merkkimäärältään kumpikin ryhmä on kuitenkin suunnilleen yhtä pitkä.

Koska tietosuoja-asetus määrää, että tietosuojaselosteiden on oltava saatavilla rekisterinpitäjien verkkosivuilla maksutta, aineiston kerääminen verkosta oli suhteellisen vaivatonta, eikä siihen tarvittu erillisiä lupia. Tutkimuseettisesta näkökulmastakin aineiston kerääminen ja käyttäminen tutkimuksessa on varsin ongelmattonta, sillä materiaali ei ole mitenkään arkaluontoista ja se on verkossa saatavilla kenelle tahansa. Hankalampaa aineiston valinnassa oli sen määrittäminen, mikä teksti on käännös (ja mistä kielestä) ja mikä taas kirjoitettu alun perin suomeksi – tätä tietoa kun ei käsillä olevan tyyppisissä teksteissä yleensä mainita. Siispä on tyydyttävä olettamukseen, että suomalaisen, Suomessa ja pääsääntöisesti suomen kielellä toimivan tahon tietosuojaseloste on mitä luultavimmin laadittu alun perin suomeksi, kun taas suuremman, yhdysvaltalaisen mutta Suomessa toimintaa harjoittavan yhtiön suomenkielinen tietosuojaseloste on mitä luultavimmin käännös englannista. On kuitenkin pidettävä mielessä, että esimerkiksi käännöksen lähtökielestä ei tämän tutkimuksen puitteissa ole mahdollista saada täyttä varmuutta. Myöskään rinnakkaistekstien osalta ei voida päästä täyteen varmuuteen siitä, etteikö esimerkiksi tekstin jokin osa voisi olla käännös, joskin tämä lienee epätodennäköistä. Tällaisia mahdollisuuksia ja niiden vaikutusta analyysin tuloksiin ja luotettavuuteen pyrittiin minimoimaan keräämällä tarpeeksi laaja aineisto useista erilaisista lähteistä.

4.3 Tutkimusmenetelmä

Isosävin ja Lappalaisen (2015: 14) mukaan puhuttelua voidaan tutkia joko aitojen vuorovaikutustilanteiden tai asenteiden ja käsitysten näkökulmista. Tässä tutkimuksessa keskitytään tämä jaottelun mukaan aitoihin vuorovaikutustilanteisiin eli aitoihin esimerkkeihin puhuttelusta kirjallisessa muodossa. Tutkimusaineiston tekstit edustavat sellaista tekstilajia, jonka olennaisessa osassa on osoittaa, keitä tekstissä kuvatut toimenpiteet ja oikeudet koskevat. Suora puhuttelu on yksi tapa osoittaa tämä (VISK, määritelmät, s.v. *puhuttelu*). Tässä tutkimuksessa puhuttelua tutkitaan vertailevalla otteella aineistosta, joka sisältää niin käännöksiä kuin rinnakkaistekstejäkin (ks. alaluku 4.2 edellä). Tutkimuksessa yhdistyvät myös määrälliset ja laadulliset menetelmät, sillä aineistosta tarkastellaan niin puhuttelumuotojen esiintymistiheyttä kuin myös tilanteita, joissa puhuttelua käytetään.

Tutkimuksessa sovelletaan Eveliina Tolvasen (2016) käyttämää luokittelua, joka jakaa teksteissä esiintyvät puhuttelumuodot seuraaviin luokkiin: (1) toisen persoonan verbinmuodot (2) persoonapronominit (3) omistusta ilmaisevat pronominit (4) possessiivisuffiksit ja (5) imperatiivit. Toisen persoonan verbinmuotoihin luetaan sekä yksikön että monikon toista persoonaa ilmaisevat verbin finiittimuodot, joilla lukijaa puhutellaan suoraan (esimerkiksi *luet* tai *käytät*). Persoonapronomineihin sisältyvät pronominit *sinä* ja *te*, omistusta ilmaiseviin pronomineihin edellä mainittujen persoonapronominien genetiivimuodot *sinun* ja *teidän*, possessiivisuffikseihin niin substantiiveihin kuin verbien infiniittimuotoihin liittyvät possessiivisuffiksit (esimerkiksi *henkilötietosi* tai *käyttäessäsi*) ja imperatiiveihin verbien käskymuodot (esimerkiksi *lue*). Koska käsillä olevan tutkimuksen aineistossa esiintyi vain yksittäisiä omistusta ilmaisevia pronomineja, yllä olevan jaottelun luokat 2 ja 3 on tämän tutkielman analyysissa yhdistetty. Tutkimusaineistossa ei myöskään esiintynyt lainkaan teitittelyä, joten käytännössä persoonapronominien ja omistusta ilmaisevien pronominin luokkaan kuuluu ainoastaan *sinä*-pronomini ja sen eri taivutusmuodot ja toisen persoonan verbinmuotoihin yksikön toista persoonaa ilmaisevat muodot. Näin ollen tässä tutkimuksessa hyödynnettävä lopullinen luokittelu on seuraavanlainen:

- (1) yksikön toista persoonaa ilmaisevat verbinmuodot
- (2) *sinä*-pronomini
- (3) possessiivisuffiksit ja
- (4) imperatiivit.

Aineiston rajaamisen jälkeen kukin analysoitavaksi valittu teksti käytiin läpi sana sanalta yllä kuvattua, hieman muokattua Tolvasen (2016) jaottelua hyödyntäen. Tekstit käytiin läpi manuaalisesti Microsoft Word -ohjelmassa, jossa jokaisesta tekstistä poimittiin yllä esitetyn jaottelun mukaiset puhuttelumuodot, minkä jälkeen laskettiin niiden lukumääräinen ja prosentuaalinen osuus tekstistä. Jokaisesta yksittäisestä tekstikatkelmasta laskettiin seuraavat luvut:

- (1) puhuttelumuotoesiintymien kokonaislukumäärä
- (2) kunkin puhuttelumuotoluokan esiintymien erillinen lukumäärä sekä
- (3) puhuttelumuotoesiintymien kokonaislukumäärän prosentuaalinen osuus koko tekstistä.

Lopuksi samat luvut laskettiin vielä kummankin aineiston eli käännösten ja rinnakkaistekstien osalta kokonaisuutena. Lisäksi kummankin aineiston osalta laskettiin kunkin puhuttelumuotoluokan prosentuaalinen osuus kyseisen aineiston puhuttelumuotoesiintymien kokonaislukumäärästä. Vertailukelpoisten lukujen saamiseksi laskettiin myös kunkin puhuttelumuotoluokan esiintymien lukumäärä tuhatta sanaa kohden kunkin aineiston osalta. Lopuksi katsottiin vielä yleisimpiä sanoja (*sinä*-pronominin osalta taivutusmuotoja), joihin suora puhuttelu aineistossa liittyy ja laskettiin kummankin aineiston osalta kaikkien neljän puhuttelumuotoluokan kolme yleisintä puhuttelua sisältävää sanaa tai muotoa.

Saadut luvut ja prosenttiluvut jäsenneltiin havainnolliseen muotoon taulukoihin, minkä jälkeen lukujen ja prosenttien vertailu oli vaivatonta. Taulukot sekä taulukoiden ja analyysin perusteella tehdyt johtopäätökset esitetään seuraavaksi luvussa 5.

5 Analyysi ja tulokset

Tässä luvussa esitellään aineiston analyysi ja analyysin tulokset. Aineistoa tarkastellaan ensin alaluvussa 5.1 yhtenä kokonaisuutena, minkä jälkeen alaluvuissa 5.2 ja 5.3 paneudutaan tarkemmin käännöksiin ja rinnakkaisteksteihin erillisinä aineistoina ja alaluvussa 5.4 esitellään aineiston analyysin tulosten perusteella tehdyt päätelmät.

5.1 Yleiskatsaus tutkimusaineiston analyysiin

Tässä alaluvussa tarkastellaan koko tutkimusaineistoa eli sekä käännöksistä että rinnakkaisteksteistä poimittuja osioita. Taulukossa 1 eritellään tutkimusaineiston laajuutta tärkeimpien lukujen valossa, minkä jälkeen siirrytään tarkastelemaan analyysin tuloksia tarkemmin:

	Käännökset	Rinnakkaistekstit
Tekstien lukumäärä	6	9
Tekstien sanamäärä yhteensä	5 142	4 707
Tekstien sanamäärä, keskiarvo	857	523
Tekstien merkkimäärä yhteensä	48 582	47 476
Tekstien merkkimäärä, keskiarvo	8 097	5 275

Taulukko 1. Käännös- ja rinnakkaistekstiaineistojen laajuudet.

Yllä oleva taulukko 1 kuvaa aineiston laajuuden käännösten ja rinnakkaistekstien osalta sanoissa ja merkeissä laskettuna. Taulukosta nähdään, että käännökset ovat keskimäärin rinnakkaistekstejä pidempiä, mutta aineiston kokonaislaajuus on käännösten ja rinnakkaistekstien osalta suurin piirtein yhtä suuri, sillä rinnakkaistekstejä on lukumääräisesti enemmän kuin käännöksiä. Seuraavaksi tarkastellaan puhuttelumuotojen esiintymistä yllä eritellyissä aineistoissa.

Tutkimusta aloitettaessa oli hypoteesina, että lukijan suora puhuttelu olisi yleisempää käännöksissä kuin rinnakkaisteksteissä (ks. luku 1), sillä aiemmassa suomenkielisten käännösten kieltä koskevassa tutkimuksessa on saatu tämänsuuntaisia tuloksia ainakin persoonapronominien käytön osalta (esimerkiksi Mauranen & Tiittula 2005; Tirkkonen-Condit

2002). Alla taulukossa 2 on eritelty puhuttelumuotojen kokonaismäärä käännösaineistossa ja rinnakkaistekstiaineistossa:

	Käännökset	Rinnakkaistekstit
Sanamäärä yhteensä	5 142	4 707
Puhuttelumuotojen määrä yhteensä	595	148
Puhuttelumuotojen %-osuus kokonaissanamäärästä	11,6	3,1
Puhuttelumuotojen esiintymien määrä / 100 sanaa	12	3

Taulukko 2. Käännös- ja rinnakkaistekstiaineistojen laajuus sekä puhuttelumuotoesiintymien kokonaismäärä kummassakin aineistossa.

Yllä olevasta taulukosta nähdään selvästi, että puhuttelumuotoja on käännösaineistossa huomattavasti enemmän kuin rinnakkaisteksteissä. Analysoitujen käännöstekstien yhteispituus oli 5 142 sanaa, josta puhuttelumuotoja oli 595 sanaa eli noin 11,6 prosenttia. Analysoitujen rinnakkaistekstien yhteispituus taas oli 4 707 sanaa, josta puhuttelumuotoja oli 148 sanaa eli noin 3,1 prosenttia. Sataa sanaa kohden käännöksissä oli siis noin 12 puhuttelumuotoa ja rinnakkaisteksteissä 3, eli käännöksissä oli lähes nelinkertainen määrä puhuttelumuotoja rinnakkaisteksteihin verrattuna. Tulos tukee esimerkiksi Maurasen ja Tiittulan (2005: 36, 66–67) huomioita siitä, että käännössuomessa esiintyy enemmän persoonapronomineja kuin alkuperäissuomessa. Alle esimerkkiin 1 on poimittu sekä käännös- että rinnakkaistekstiaineistosta suunnilleen samaa asiaa ilmaiseva virke, joista toisessa on käytetty suoraa puhuttelua ja toisessa ei. Käännös on merkitty tunnuksella (K) ja rinnakkaisteksti tunnuksella (R), suora puhuttelu on kursivoitu ja katkelman lopussa suluissa on ilmoitettu teksti, josta esimerkki on poimittu:

- (1) (K) *Sinulla* on oikeus tarkastaa, mitä *sinua* koskevia tietoja rekisteriin on tallennettu. (K5)
- (R) Henkilöllä on oikeus tarkastaa, mitä tietoja hänestä on talletettu asiakasrekisteriin. (R4)

Esimerkistä nähdään, kuinka käännöksessä lukijaa puhutellaan suoraan, mutta rinnakkaistekstissä lukijaan viitataan sanalla *henkilö*. Esimerkin puhuttelumuodot edustavat aiemmin esitellyn luokittelun (ks. alaluku 4.3) luokkaa 2 eli *sinä*-pronominia. Vertailun vuoksi

alle esimerkkiin 2 on poimittu kummastakin aineistosta kaksi suunnilleen samaa asiaa ilmaisevaa virkettä, joista yhteensä kolmessa esiintyy lukijan suoraa puhuttelua ja yhdessä ei:

- (2) (K a.) Mikäli *olet* kuitenkin tyytymätön *henkilötietojesi* käsittelyyn, *sinulla* on oikeus saattaa asia tietosuojaviranomaisen tutkittavaksi. (K5)
- (K b.) – – *sinulla* on oikeus valittaa tietosuojaviranomaisille *henkilötietojesi* keräämisestämme ja käytöstämme. (K2)
- (R a.) Jos *huomaat*, että käsittelyssä on puutteita tai se on lainvastaista, *sinulla* on oikeus tehdä asiasta valitus tietosuojaviranomaiselle. (R5)
- (R b.) Henkilöllä on oikeus tehdä valitus toimivaltaiselle valvontaviranomaiselle, jos hän katsoo, että [– –] ei ole noudattanut soveltuvaa tietosuojalainsäätelyä. (R4)

Esimerkin neljästä virkkeestä kolmessa esiintyy suoraa puhuttelua. Virkkeet K a. ja R a. ovat puhuttelun osalta melko samanlaiset: kummassakin esiintyy toisen persoonan verbinmuoto (*olet* ja *huomaat*) sekä adessiivimuotoinen *sinä*-pronomini. Lisäksi virkkeessä K a. on substantiiviin *henkilötiedot* liittyvä possessiivisuffiksi (*henkilötietosi*). Virkkeet K b. ja R b. taas ovat muihin esimerkkivirkkeisiin ja toisiinsa verrattuna melko erilaiset. Virkkeessä K b. huomio kiinnittyy lukijan suoran puhuttelun ohella myös monikon ensimmäisen persoonan muotoihin, joilla viitataan tietosuojaselosteen laatijaan (*keräämisestämme* ja *käytöstämme*). Kun virkettä verrataan virkkeeseen R b., jossa ei esiinny lainkaan persoonaviittauksia, ero on varsin huomattava. Nämä neljä esimerkkivirkettä osoittavat sen, että tekstikohtaiset erot käännösten ja rinnakkaistekstien välillä voivat olla huomattavia tai vaihtoehtoisesti hyvin pieniä. Myös aineistojen sisällä näyttäisi olevan jonkin verran hajontaa (ks. tarkemmin alaluvut 5.2 ja 5.3 jäljempänä).

Alla taulukossa 3 on edelleen eritelty kunkin alaluvussa 4.3 esitettyä luokittelua noudattavan puhuttelumuotoluokan osuudet käännöksissä ja rinnakkaisteksteissä erikseen:

	Käännökset, esiintymiä (%)	Rinnakkaistekstit, esiintymiä (%)
2. pers. verbinmuoto	181 (30)	27 (18)
<i>sinä</i> -pronomini	91 (15)	35 (24)
Possessiivisuffiksi	308 (52)	80 (54)
Imperatiivi	15 (3)	6 (4)
Yhteensä	595 (100)	148 (100)

Taulukko 3. Kunkin puhuttelumuotoluokan esiintymien lukumäärä kussakin aineistossa sekä kunkin puhuttelumuotoluokan prosentuaalinen osuus kyseisen aineiston puhuttelumuotoesiintymien kokonaismäärästä.

Taulukosta 3 voi nähdä heti ensimmäisellä silmäyksellä sen, mitä oli nähtävissä jo taulukossa 2 – rinnakkaisteksteissä esiintyy lukumääräisesti huomattavasti vähemmän puhuttelumuotoja kuin käännöksissä. Lukumäärien suureen eroon vaikuttavat toki osaltaan erot aineistojen laajuuksissa eli se, että käännösaineisto oli laajuudeltaan noin 400 sanaa laajempi kuin rinnakkaistekstiaineisto. Tämä ei kuitenkaan yksin riitä selittämään puhuttelumuotojen lukumäärien eroja, vaan kuten taulukosta 2 voidaan nähdä, ero on prosentuaalisestikin tarkasteltuna varsin selkeä. Taulukossa 3 yllä on lisäksi kuvattu kunkin puhuttelumuotoluokan prosentuaaliset osuudet kyseisen aineiston puhuttelumuotojen kokonaismäärästä sekä käännösten että rinnakkaistekstien osalta.

Taulukon 3 lukujen perusteella voidaan todeta, että kummassakin aineistossa imperatiivia esiintyi puhuttelumuodoista vähiten, kun taas possessiivisuffiksi oli molemmissa aineistoissa ylivoimaisesti yleisin. On mielenkiintoista huomata, että vaikka kokonaisuutena tarkasteltuna käännöksissä esiintyi selvästi enemmän puhuttelua kuin rinnakkaisteksteissä, imperatiivin ja possessiivisuffiksin prosentuaalinen osuus on kummankin aineiston puhuttelumuodoista lähes täsmälleen sama – käännöksissä imperatiivi vastasi 3:a prosenttia puhuttelumuodoista ja possessiivisuffiksi 52:ta prosenttia, rinnakkaisteksteissä vastaavat luvut olivat 4 ja 54. Sen sijaan toiseksi ja kolmanneksi yleisimpien puhuttelumuotojen osalta käännökset ja rinnakkaistekstit olivat toistensa vastakohdat. Käännöksissä toiseksi yleisin puhuttelumuoto oli toisen persoonan verbinmuoto, joita oli 30 prosenttia kyseisen aineiston puhuttelumuodoista. Kolmanneksi yleisin muoto oli *sinä*-pronomini, joita oli 15 prosenttia kyseisen aineiston puhuttelumuodoista. Rinnakkaistekstien osalta näiden luokkien yleisyys on siis käänteinen, eli toiseksi yleisin puhuttelumuoto oli *sinä*-pronomini, joita oli 24 prosenttia kyseisen aineiston

puhuttelumuodoista, ja kolmanneksi yleisin oli toisen persoonan verbinmuoto, joita oli 18 prosenttia kyseisen aineiston puhuttelumuodoista. Suoran puhuttelun esiintymistiheyden lisäksi näyttäisi siis siltä, että käännösten ja rinnakkaistekstien välillä on jonkin verran eroa myös siinä, millaisia suoran puhuttelun muotoja suositaan.

Pelkkiä prosenttilukuja vertailtaessa on helppo erehtyä ajattelemaan, että vertailtavat aineistot eli käännökset ja rinnakkaistekstit olisivat melko tasavertaisia keskenään, sillä kuten edellä on todettu, prosenttiluvut puhuttelumuotojen jakautumisesta kunkin aineiston sisällä ovat melko tasaiset, kun aineistoja verrataan toisiinsa (ks. taulukko 3). Toisaalta taas pelkät lukumäärät eivät ole keskenään vertailukelpoisia sellaisenaan aineistojen erilaisten laajuuksien takia. Näin ollen alle taulukkoon 4 on laskettu kunkin puhuttelumuotoluokan esiintymien lukumäärä tuhatta sanaa kohti sekä käännösten että rinnakkaistekstien osalta:

	Käännökset, esiintymiä / 1 000 sanaa	Rinnakkaistekstit, esiintymiä / 1 000 sanaa
2. pers. verbinmuoto	35	6
<i>sinä</i> -pronomini	18	7
Possessiivisuffiksi	60	17
Imperatiivi	3	1
Yhteensä	116	31

Taulukko 4. Puhuttelumuotoluokkien esiintymät tuhatta sanaa kohti käännöksissä ja rinnakkaisteksteissä. Luvut on pyöristetty lähimpään kokonaiseen lukuun.

Taulukosta nähdään siis vertailukelpoiset luvut kunkin puhuttelumuotoluokan osalta käännöksissä ja rinnakkaisteksteissä. Viimeistään näiden lukujen myötä on selvää, että puhuttelua esiintyy käännöksissä moninkertaisesti verrattuna rinnakkaisteksteihin. Ainoastaan imperatiivin osalta eron voisi sanoa olevan pienehkö. Kuten taulukosta käy selväksi, rinnakkaistekstien osalta possessiivisuffiksia lukuun ottamatta kaikkia muita puhuttelumuotoja esiintyy tuhatta sanaa kohti alle kymmenen, possessiivisuffiksejakin ainoastaan 17. Ero käännösten ja rinnakkaistekstien välillä on huomattavin possessiivisuffiksien ja toisen persoonan verbinmuotojen osalta. Käännöksissä tuhatta sanaa kohti oli 60 possessiivisuffiksia ja 35 toisen persoonan verbinmuotoa, kun rinnakkaisteksteissä vastaavat luvut olivat 17 ja 6.

Alle esimerkkiin 3 on poimittu käännöksistä ja rinnakkaisteksteistä samantyyppistä seikkaa, tässä tapauksessa kerättyjen tietojen tyyppiä, ilmaisevat katkelmat konkretisoimaan taulukosta

4 nähtävää eroa puhuttelumuotojen määrissä. Esimerkissä käännös on merkitty kirjaimella K ja rinnakkaisteksti kirjaimella R, suoraa puhuttelua sisältävä sana on kursivoitu ja mahdollinen puhuttelun osoittava liite tai pääte alleviivattu ja katkelman lopussa suluissa on ilmoitettu teksti, josta esimerkki on peräisin:

- (3) (K) *Sinun käyttösi*. Keräämme tietoa siitä, miten *käytät* tuotteitamme, kuten millaisen sisällön kanssa *toimit* ja mitä sisältöä *katsot*, millaisia ominaisuuksia *käytät*, mitä toimintoja *suoritat*, keiden ihmisten ja käyttäjätilien kanssa *olet* vuorovaikutuksessa, kuinka usein *toimit* ja mikä on *toimiesi* kesto. Kirjaamme esimerkiksi lokiin kun *käytät* tai *olet* viimeksi käyttänyt tuotteitamme sekä mitä julkaisuja, videoita ja muuta sisältöä *katselet* tuotteissamme. (K4)

(R) Keräämme käyttäjästä tämän tietosuojalausekkeen kohdassa 2 kuvattujen käyttötarkoitusten kannalta tarpeellisia henkilötietoja. [– –]¹ Sivu, jolta käyttäjä on siirtynyt sivuillemme, laitteen malli, yksilöllinen laite- ja/tai eväsetunniste, tiedonkeruun kanava (internetselain, mobiiliselain, sovellus), selaimen versio, IP-osoite, istuntotunniste, istunnon aika ja kesto sekä näytön tarkkuus ja käyttöjärjestelmä. (R2)

Yllä olevan esimerkin tekstit on poimittu sillä perusteella, että ne ilmaisevat samantyyppistä asiaa eli kerättyjen tietojen tyyppiä. Koska yhtiöiden keräämien tietojen välillä on eroja, täysin samaa asiaa ilmaisevia kohtia on mahdotonta löytää. Esimerkistä kuitenkin nähdään, että molemmissa teksteissä selostetaan sitä, millaisia tietoja lukijasta kerätään. Käännöksessä puhuttelua esiintyy hyvin runsaasti – katkelman 57 sanassa esiintyy puhuttelua 13 kertaa. Kyseisessä katkelmassa hallitsevana puhuttelumuotona on toisen persoonan verbinmuoto, mutta myös possessiivisuffiksia ja persoonapronominia on käytetty. Suomalaiselle lukijalle erityisesti heti katkelman alussa esiintyvä rakenne *sinun käyttösi* (persoonapronomini + possessiivisuffiksi) tuntuu oudolta, sillä suomen kielessä on puhuttelutilanteessa tavanmukaisempaa jättää pronomini pois ja osoittaa vastaanottaja ainoastaan possessiivisuffiksilla (Lappalainen 2015: 82). Esimerkin rinnakkaistekstissä taas suoraa puhuttelua ei esiinny lainkaan, vaan lukijaan viitataan epäsuorasti sanalla *käyttäjä*. Tällainen epäsuora viittaaminen oli rinnakkaisteksteille ominaista, vaikka teksteissä esiintyi myös suoraa puhuttelua (ks. tarkemmin alaluku 5.3 jäljempänä).

Edellä taulukossa 4 kuvattuja lukuja on myös mielenkiintoista verrata Tolvasen (2016) lukuihin – Tolvasen tutkimus käsitteli suomen- ja ruotsinkielisiä viranomaistekstejä, ja aineisto oli poimittu Suomen Kansaneläkelaitoksen ja Ruotsin Eläkeviraston teksteistä (Tolvanen 2016: 5). On hyvä huomata, että vaikka osa Tolvasen teksteistä oli käännöksiä, hän ei keskittynyt tutkimuksessaan kääntämiseen, eikä tehnyt aineistossaan varsinaista eroa käännösten ja ei-käännettyjen tekstien välille. Hänen Kelan teksteistä kokoamansa suomenkielinen aineisto

¹ Vertailukelpoisen muodon saavuttamiseksi välistä on poistettu luettelomaisen osuuden väliotsikko.

koostui kuitenkin alkuperäisistä suomenkielisistä teksteistä, eli tästä aineistosta saadut tulokset ovat tässä mielessä vertailukelpoisia tutkimukseni rinnakkaistekstiaineiston kanssa. Vertailun helpottamiseksi olen koonnut alla olevaan taulukkoon keskimmäiseen sarakkeeseen Tolvasen Kela-aineiston luvut (mas. 13; taulukkoa varten on yhdistetty persoonapronominien ja omistusta ilmaisevien pronomien luokat) ja oikeanpuolimmaiseen tämän tutkimuksen rinnakkaistekstiaineiston luvut (esiintymiä / 1 000 sanaa):

	Tolvanen	Tämä tutkimus
2. pers. verbinmuoto	40,1	6
Persoon- ja omistusta ilmaisevat pronominit	7,1	7
Possessiivisuffiksi	19,7	17
Imperatiivi	7,2	1
Yhteensä	74,1	31

Taulukko 5. Eveliina Tolvasen (2016: 13) suomenkielisen Kela-aineiston ja tämän tutkimuksen rinnakkaistekstiaineiston sisältämät puhuttelumuodot laskettuna tuhatta sanaa kohti.

Lukumääräisesti Tolvasen aineistossa ja tämän tutkimuksen aineistossa persoonapronomineja ja omistusta ilmaisevia pronomineja oli saman verran eli noin 7 esiintymää tuhatta sanaa kohti. Possessiivisuffiksienkin määrä on melko samanlainen, Tolvasen aineistossa 19,7 ja tämän tutkimuksen aineistossa 17 esiintymää tuhatta sanaa kohti. Imperatiivin osalta ero on noin 6 esiintymää, mutta suurin ja melko huomattavakin ero on toisen persoonan verbinmuotojen määrässä. Tolvasen aineistossa toisen persoonan verbinmuotoja esiintyi 40,1 kertaa tuhatta sanaa kohti, kun tämän tutkimuksen aineistossa vastaava luku on 6. Ero on valtava, ja jopa suurempi kuin tämän tutkimuksen käännösten ja rinnakkaistekstien välillä (ks. taulukko 4 edellä). Eron voisi tietenkin ajatella johtuvan esimerkiksi tekstilajien välisistä eroista, mutta luultavasti sille löytyy vielä tätäkin loogisempi selittäjä. Kuten aiemmin alaluvuissa 2.1 ja 2.2 mainittiin, Kela teki vuonna 2011 päätöksen siirtyä sinutteluun suurimmassa osassa viestintäänsä (Oikeusasiamies 2014). On siis hyvin todennäköistä, että tästä syystä Tolvasen aineistossa lukijaa sinutellaan systemaattisesti, kun taas tämän tutkimuksen aineistossa hajontaa löytyy jo senkin vuoksi, että tekstit eivät ole yhdeltä taholta peräisin.

Kahdessa seuraavassa alaluvussa siirrytään tarkastelemaan käännös- ja rinnakkaistekstiaineistoissa esiintynyttä puhuttelua yksityiskohtaisemmin.

5.2 Lukijan puhuttelu käännöksissä

Tässä alaluvussa keskitytään lähemmin puhuttelumuotojen esiintymiseen käännösaineistossa. Käännösaineiston muodostavat yhdysvaltalaisen yhtiöiden Applen, Netflixin, Microsoftin, Facebookin, McDonald'sin ja Electronic Artsin tietosuojaselosteet. Tämän alaluvun sisältämissä taulukoissa ja taulukoiden selitteissä näihin viitataan kirjain–numero-yhdistelmillä K1–K6 edellä luetellussa järjestyksessä paitsi tilan säästämiseksi, myös siksi, ettei tutkimuksen ja analyysin kannalta ole relevanttia liittää tuloksia tiettyyn yhtiöön.

Alaluvussa 5.1 tarkasteltiin puhuttelumuotojen esiintymien lukumäärää ja prosentuaalista osuutta käännösaineistossa yhteensä (ks. taulukot 2–4 edellä), joten tässä alaluvussa siirrytään tutkimaan muun muassa sitä, miten puhuttelumuodot jakautuvat eri käännöksissä. Alla taulukossa 6 esitetään kunkin käännöksen pituus sanoina ja merkkeinä, kunkin käännöksen sisältämien puhuttelumuotoesiintymien lukumäärä ja puhuttelumuotoesiintymien prosentuaalinen osuus koko käännöksen sanamäärästä:

	K1	K2	K3	K4	K5	K6
Sanamäärä	551	842	387	1435	880	1047
Merkkimäärä	5 298	8 205	3 701	13 084	8 426	9 868
Puhuttelumuodot, esiintymiä	55	91	32	156	85	175
Puhuttelumuodot, %-osuus	10,0	10,9	8,3	10,9	9,7	16,7

Taulukko 6. Käännösten laajuus sekä puhuttelumuotoesiintymät kussakin käännöksessä.

Taulukosta 6 nähdään, että aineiston sisältämien käännösten pituuksissa ja täten myös puhuttelumuotojen esiintymien lukumäärissä on joiltakin osin jopa huomattavia eroja, mutta neljässä tekstissä kuudesta puhuttelumuotojen prosentuaalinen osuus on hyvin lähellä 10:tä prosenttia. Tästä poikkeaa hieman kaksi tekstiä – toisessa näistä teksteistä puhuttelun prosentuaalinen osuus on noin 8,3 prosenttia ja toisessa noin 16,7. Suurin osa teksteistä on siis myös yksittäisen tekstin tasolla hyvin lähellä koko käännösaineistosta laskettua puhuttelun prosentuaalista osuutta (11,6 %, ks. taulukko 2 alaluvussa 5.1).

Seuraavaksi tarkastellaan puhuttelumuotojen esiintymistä kussakin käännöksessä yksityiskohtaisemmin. Alle taulukkoon 7 on koottu kunkin puhuttelumuotoluokan esiintymien määrä kussakin käännöksessä:

	K1	K2	K3	K4	K5	K6
2. pers. verbinmuoto	18	23	11	61	31	37
<i>sinä</i> -pronomini	5	15	8	28	16	18
Possessiivisuffiksi	30	50	12	62	37	117
Imperatiivi	2	3	1	5	1	3
Yhteensä	55	91	32	156	85	175

Taulukko 7. Puhuttelumuotoluokkien esiintymät kussakin käännöksessä.

Yllä olevan taulukon tiedot tukevat alaluvussa 5.1 tarkasteltuja tietoja käännösaineistosta kokonaisuutena. Käännösaineiston yleisin puhuttelumuoto oli possessiivisuffiksi, toiseksi yleisin toisen persoonan verbinmuoto, kolmanneksi yleisin *sinä*-pronomini ja vähiten esiintynyt imperatiivi. Kun tarkastellaan kutakin käännöstä yksittäisenä tekstinä, huomataan, että jokainen käännös noudattaa samaa jakaumaa. On kuitenkin hyvä huomata, että esimerkiksi käännöksissä K3 ja K4 ero possessiivisuffiksin ja toisen persoonan verbinmuodon välillä on hyvin pieni. Jotta käännöksiä olisi vielä helpompi vertailla keskenään, alle taulukkoon 8 on laskettu kunkin puhuttelumuotoluokan esiintymien lukumäärä kussakin tekstissä tuhatta sanaa kohti:

	K1	K2	K3	K4	K5	K6
2. pers. verbinmuoto / 1 000 sanaa	33	27	28	42	35	35
<i>sinä</i> -pronomini / 1 000 sanaa	9	18	21	19	18	17
Poss.suff. / 1 000 sanaa	54	59	31	43	42	112
Imperatiivi / 1 000 sanaa	4	4	3	3	1	3
Yhteensä	100	108	83	107	96	167

Taulukko 8. Puhuttelumuotoluokkien esiintymät tuhatta sanaa kohti kussakin käännöksessä. Luvut on pyöristetty lähimpään kokonaiseen lukuun.

Seuraavissa esimerkeissä 4–7 tarkastellaan taulukoissa 7 ja 8 kuvattuja puhuttelumuotoluokkia käännöksistä poimittujen aitojen katkelmien avulla järjestyksessä yleisimmästä harvinaisimpaan. Kussakin esimerkissä tarkasteltavan puhuttelumuodon sisältävä sana on kursivoitu ja mahdollinen puhuttelun osoittava liite tai päätte alleviivattu. Muut katkelmassa mahdollisesti esiintyneet puhuttelumuodot on jätetty esimerkeissä merkitsemättä, jotta huomio

kiinnittyy kulloinkin tarkasteltavaan muotoon. Ensimmäisenä esimerkissä 4 tarkastellaan possessiivisuffiksia, joka on siis yleisin käännösaineistossa esiintynyt puhuttelumuoto:

- (4) Keräämämme tiedot riippuvat *toiminnastasi* [– –] kanssa sekä *tekemistäsi* valinnoista, mukaan lukien *tietosuoja-asetuksistasi*, sekä *käyttämistäsi* tuotteista ja ominaisuuksista. (K3)

Possessiivisuffiksia esiintyi käännösaineistossa puhuttelumuodoista eniten (60 / 1 000 sanaa), ja sitä käytettiin esimerkiksi kohdissa, joissa eritellään, mitä tietoja lukijasta kerätään ja millä perusteella. Yllä olevassa katkelmassa kerrotaan, miten lukijan toiminta (*tekemäsi valinnat, tietosuoja-asetuksesi, käyttämäsi tuotteet*) vaikuttaa hänestä kerättyihin tietoihin. Tekstissä esiintyvä puhuttelu on sinänsä suomen konventioiden mukaista, että persoonapronominin sijaan puhuttelu osoitetaan pelkällä possessiivisuffiksilla (ks. alaluku 2.2). Tämän tyyppinen puhuttelu on kuitenkin tästä huolimatta huomattavasti harvinaisempaa rinnakkaistekstiaineistossa (käännöksissä 60 / 1 000 sanaa, rinnakkaisteksteissä 17 / 1 000 sanaa), joten sen voitaneen olettaa olevan ainakin osittain lähtökielen eli englannin vaikutuksen seurausta eli *interferenssiä* (interferenssistä ks. alaluku 3.2).

Seuraavaksi esimerkissä 5 tarkastellaan käännösaineistossa toiseksi eniten esiintynyttä puhuttelumuotoa, toisen persoonan verbinmuotoa:

- (5) a. Keräämämme tiedot riippuvat siitä, mitä tuotteita *käytät* ja miten niitä *käytät*. (K6)
b. Jos *olet* antanut meille luvan lähettää sinulle markkinointiviestintää, *voit* myöhemmin kieltäytyä viestinnästä viesteihin sisällyttämämme ohjeiden mukaisesti. (K6)

Esimerkki 5 esittää kaksi esimerkkitilannetta, jollaisessa toisen persoonan verbinmuotoa esiintyi käännöksissä. Toisen persoonan verbinmuotoa esiintyi käännösaineistossa 35 kertaa tuhatta sanaa kohti ja muun muassa samantyyppisissä kohdissa kuin possessiivisuffiksiakin (ks. esimerkki 4 edellä) eli kohdissa, joissa kuvattiin, mitä lukijan tietoja kerätään ja millä perusteella. Tällaista esiintymää edustaa esimerkki 5 a., jossa kerrotaan, että lukijan käyttämät tuotteet vaikuttavat siihen, mitä tietoja hänestä kerätään. Toinen tilanne, jossa toisen persoonan verbiä muun muassa esiintyi, oli lukijan oikeuksien kuvailu. Tällaista tilannetta edustaa virke 5 b., jossa kerrotaan lukijan oikeudesta kieltäytyä markkinointiviestinnästä. Kuten todettiin myös esimerkin 4 osalta, tässäkin esimerkissä esiintyvä puhuttelu noudattaa suomen kielen puhuttelukonventioita siinä mielessä, että puhuttelu ilmaistaan pelkällä verbin taivutuksella ilman persoonapronominia (ks. alaluku 2.2). Myös tässä kohtaa on kuitenkin todettava, että rinnakkaistekstiaineistossa tällainenkin puhuttelu oli huomattavasti harvinaisempaa (käännöksissä 35 / 1 000 sanaa, rinnakkaisteksteissä 6 / 1 000 sanaa). Lähtötekstin vaikutus (ks.

alaluvut 3.1 ja 3.2) voisi siis olla toisen persoonan verbinmuotojenkin runsaan käytön yksi mahdollinen selittävä tekijä.

Alla esimerkissä 6 tarkastellaan kolmanneksi eniten, ja samalla toiseksi vähiten, käännösaineistossa esiintynyttä puhuttelumuotoa, *sinä*-pronominia:

- (6)
- a. Lisäksi *sinulla* on tietyissä tilanteissa oikeus pyytää *sinua* koskevien henkilötietojen poistamista rekisteristä sekä pyytää niiden siirtämistä toiselle yritykselle. (K5)
 - b. Jotta voimme tarjota *sinulle* [– –]tuotteet, meidän täytyy käsitellä *sinua* koskevia tietoja. (K4)
 - c. *Sinä* hallitset kaikkia henkilötietoja, jotka annat meille verkkopalveluidemme kautta. (K5)

Sinä-pronominin käyttö on suomessa huomattavasti harvinaisempaa kuin englannissa, jossa *you*-pronominia käytetään neutraalina yleisilmaisuna, kun puhutaan ihmisistä yleensä (Isosävi & Lappalainen 2015: 12). Käännösaineistossa *sinä*-pronominia kuitenkin esiintyi jonkin verran (n. 15 % kaikista puhuttelumuodoista, ks. tarkemmin taulukot 3 ja 4), ja sitä käytettiin esimerkiksi siinä yhteydessä, kun lukijalle kerrottiin tämän oikeuksista (virkkeet 6 a. ja c.) tai kun lukijalle kerrottiin, mitä tietoja hänestä kerätään ja mitä varten (virke 6 b.). Voisi olettaa, että englannin kielen vaikutus näkyisi juuri tämän puhuttelumuodon käytössä, sillä kuten mainittua, englannissa *you*-pronominia käytetään paljon runsaammin kuin suomen kielessä *sinä*-pronominia. *Sinä*-pronominia toki esiintyi käännöksissä 18 kertaa tuhatta sanaa kohti ja rinnakkaisteksteissä vain 7, mutta tästä huolimatta rinnakkaistekstiaineistossa *sinä*-pronomini oli suhteessa yleisempi puhuttelumuoto kuin käännöksissä (käännöksissä 15 % kaikista puhuttelumuodoista, rinnakkaisteksteissä 24 %).

Seuraavaksi esimerkissä 7 tarkastellaan käännösaineistossa vähiten esiintynyttä puhuttelumuotoa, imperatiivia:

- (7) *Käytä* hieman aikaa tietosuojakäytäntöihimme tutustumiseen ja *ota* meihin yhteyttä, jos *sinulla* on kysyttävää. (K1)

Imperatiivi esiintyi käännösaineistossa ainoastaan kolme kertaa tuhatta sanaa kohti. Sitä esiintyi siis vain vähän verrattuna toisen persoonan verbinmuotoihin (35 / 1 000 sanaa), possessiivisuffikseihin (60 / 1 000 sanaa) ja *sinä*-pronominiin (18 / 1 000 sanaa) (ks. esimerkiksi taulukko 4 edellä), mutta kuitenkin jokaisessa käännösaineiston tekstissä vähintään kerran. Imperatiivi esiintyi esimerkiksi yhteyksissä, joissa pyydettiin lukijaa perehtymään johonkin

tekstiin tai ottamaan tarvittaessa yhteyttä esimerkiksi tietosuojaselosteen laatijaan, kuten yllä olevassakin esimerkissä.

Jotta aineistosta ja siinä esiintyvistä suorasta puhuttelusta saataisiin mahdollisimman konkreettinen käsitys, taulukkoon 9 alla on koottu jokaisen puhuttelumuotoluokan osalta kolme yleisintä sanaa, johon puhuttelumuoto liittyy (*sinä*-pronominin osalta kolme yleisintä taivutusmuotoa):

	2. persoonan verbinmuoto (esiintymiä)	<i>sinä</i>-pronomini (esiintymiä)	Possessiivisuffiksi (esiintymiä)	Imperatiivi (esiintymiä)
Yleisin	<i>olla</i> (28)	<i>sinulle</i> (33)	<i>tieto</i> (25)	<i>lukea</i> (8)
2. yleisin	<i>voida</i> (26)	<i>sinusta</i> (13)	<i>antaa, käyttää</i> (11)	<i>ottaa yhteyttä</i> (5)
3. yleisin	<i>käyttää</i> (20)	<i>sinulla</i> (12)	<i>henkilötieto, tili</i> (10)	

Taulukko 9. Kolme yleisintä sanaa, johon puhuttelumuoto liittyy, kunkin puhuttelumuotoluokan osalta käännoksissä.²

Taulukosta nähdään käännoaineiston yleisimmät puhuttelua sisältävät sanamuodot ja sanat, joihin puhuttelumuodot liittyvät. Verbeistä selvästi yleisimmät olivat *olla* (esimerkiksi ”kun *olet* yhteydessä” tai ”jos *olet* kieltäytynyt markkinointisähköposteista”), *voida* (esimerkiksi ”*voit* pyytää pääsyä henkilötietoihin” tai ”*voit* vastustaa henkilötietojesi käsittelyä”) ja *käyttää* (esimerkiksi ”keräämme tietoa siitä, miten *käytät* tuotteitamme”). *Sinä*-pronomini taas esiintyi aineistossa ylivoimaisesti eniten allatiivimuodossa *sinulle* (esimerkiksi ”palvelujemme tarjoamiseksi *sinulle*”), mutta usein myös elatiivimuodossa *sinusta* (esimerkiksi ”saamme tietoja *sinusta*”) ja adessiivimuodossa *sinulla* (esimerkiksi ”*sinulla* on oikeus”). Possessiivisuffiksi liittyi useimmiten substantiivin *tieto* (esimerkiksi ”*tietojesi* käsittely”) ja usein myös esimerkiksi verbien *antaa* ja *käyttää* infiniittimuotoihin (esimerkiksi ”*antamasi* tiedot” tai ”*käyttäessäsi* tuotteitamme”) sekä substantiiveihin *henkilötieto* ja *tili* (esimerkiksi ”voimme kerätä *henkilötietojasi*” tai ”*tiliäsi* koskevat tiedot”). Imperatiiveista yleisimmät olivat muodot *lue* ja *ota yhteyttä* (esimerkiksi ”*lue* lisää siitä, miten käytämme evästeitä” tai ”*ota* meihin yhteyttä, jos sinulla on kysyttävää”). Kaikki muut imperatiivimuodot esiintyivät

² Imperatiivin osalta annettu vain kaksi yleisintä sanaa, sillä kaikki muut sanat esiintyivät aineistossa vain kerran.

aineistossa vain kerran, minkä vuoksi taulukoon ei ole merkitty kolmanneksi yleisintä imperatiivia.

Edellä tässä alaluvussa kuvatun tarkastelun sekä kvantifioinnista saatujen tulosten perusteella voidaan todeta, että käännösaineiston sisältämät tekstit ovat lukijan suoran puhuttelun osalta keskenään melko samanlaisia. Puhuttelumuotoluokkien esiintymät jakautuvat jokaisessa tekstissä samalla tavalla (ks. taulukko 6), ja puhuttelua on määrällisestikin suunnilleen saman verran, noin 10 prosenttia sanamäärästä, lähes jokaisessa tekstissä. Tästä poiketen yhdessä tekstissä puhuttelu vastaa vajaata 17:ää prosenttia tekstin sanamäärästä ja yhdessä noin 8:aa. Kokonaisuutena tarkasteltuna käännösaineisto näyttäytyy suoran puhuttelun osalta tästä huolimatta melko homogeenisenä. Seuraavassa alaluvussa siirrytään tarkastelemaan lukijan puhuttelua rinnakkaistekstiaineistossa.

5.3 Lukijan puhuttelu rinnakkaisteksteissä

Rinnakkaistekstiaineiston muodostavat suomalaisten Atrian, Sanoman, HSL:n, Alkon, Verkkokauppa.comin, Olvin, Evidensian, Mustin & Mirrin sekä K-Plussan tietosuojaselosteet. Tämän alaluvun sisältämissä taulukoissa ja taulukoiden selitteissä näihin viitataan kirjainnumero-yhdistelmillä R1–R9 edellä luetellussa järjestyksessä paitsi tilan säästämiseksi, myös siksi, ettei tutkimuksen ja analyysin kannalta ole relevanttia liittää tuloksia tiettyyn yhtiöön rinnakkaistekstienkään osalta.

Alaluvussa 5.1 tarkasteltiin puhuttelumuotojen esiintymien määrää ja prosentuaalista osuutta aineistoissa yleisesti ja alaluvussa 5.2 puhuttelua käännösaineistossa tarkemmin, joten tässä alaluvussa siirrytään tutkimaan sitä, miten puhuttelumuodot jakautuvat eri rinnakkaisteksteissä. Alla taulukossa 10 esitetään kunkin rinnakkaistekstin pituus sanoina ja merkkeinä, kunkin rinnakkaistekstin sisältämien puhuttelumuotoesiintymien määrä ja puhuttelumuotoesiintymien prosentuaalinen osuus koko rinnakkaistekstin sanamäärästä:

	R1	R2	R3	R4	R5	R6	R7	R8	R9
Sanamäärä	375	597	731	385	397	186	438	911	687
Merkkimäärä	3 612	5 980	7 298	4 094	3 983	1 801	4 613	9 162	6 933
Puhuttelumuodot, esiintymiä	7	7	1	1	19	19	6	0	89
Puhuttelumuodot, %-osuus	1,9	1,2	0,2	0,3	4,8	10,2	1,4	0,0	12,9

Taulukko 10. Rinnakkaistekstien laajuus sekä puhuttelumuotoesiintymien määrä kussakin rinnakkaistekstissä.

Kuten taulukosta 10 nähdään (ks. etenkin alin rivi), rinnakkaisteksteissä on hajontaa puhuttelun osalta huomattavasti enemmän kuin käänöksissä, joissa prosenttiosuus oli 10 prosentin luokkaa lähes jokaisessa tekstissä (ks. taulukko 6). Yhdeksästä tekstistä kuudessa puhuttelumuotojen prosentuaalinen osuus tekstin kokonaissanamäärästä on alle 2 prosenttia. Rinnakkaistekstissä R5 puhuttelua on taas 4,8 prosenttia kokonaissanamäärästä, ja rinnakkaisteksteissä R6 ja R9 puhuttelua on 10,2 ja 12,9 prosenttia. Puhuttelun prosentuaalista osuutta katsottaessa rinnakkaistekstit R6 ja R9 ovat siis lähempänä käänösaineistoa kuin rinnakkaistekstiaineistoa. Käänösaineistossa kokonaisuutena puhuttelua oli 11,6 prosenttia ja rinnakkaistekstiaineistossa 3,1 prosenttia aineiston kokonaissanamäärästä. Yksittäisten rinnakkaistekstien hajonta puhuttelun osalta on varsin mielenkiintoinen tulos. Toki jonkinasteista hajontaa on odotettavissa jo esimerkiksi tekstien laatijoiden henkilökohtaisista mieltymyksistä, mutta toisaalta saman selityksen perusteella pitäisi voida olettaa hajontaa olevan samassa määrin myös käänöksissä – ovathan nekin eri tahojen laatimia ja kääntämiä. Samansuuntaisia havaintoja on kuitenkin tehnyt esimerkiksi Tirkkonen-Condit (2005: 125) suomen kielen uniikkiaineiksia tutkiessaan. Hän havaitsi tutkimuksessaan liitepartikkeleista, että alun perin suomeksi kirjoitetuissa teksteissä oli liitepartikkelien käytön osalta huomattavaa keskinäistä vaihtelua verrattuna käännettyihin teksteihin. Mahdollisia syitä tähän tulokseen pohditaan vielä tarkemmin alaluvussa 5.4.

Seuraavaksi alla olevassa taulukossa 11 tarkastellaan puhuttelumuotoluokkien esiintymiä kussakin rinnakkaistekstissä erikseen:

	R1	R2	R3	R4	R5	R6	R7	R8	R9
2. pers. verbinmuoto	2	0	1	0	4	3	1	0	16
<i>sinä</i> -pronomini	4	1	0	0	5	7	3	0	16
Possessiivisuffiksi	1	1	0	1	10	9	2	0	56
Imperatiivi	0	5	0	0	0	0	0	0	1
Yhteensä	7	7	1	1	19	19	6	0	89

Taulukko 11. Puhuttelumuotoluokkien esiintymät kussakin rinnakkaistekstissä.

Taulukkoa 11 tarkasteltaessa huomataan, että toisin kuin käänöksissä esiintyvä puhuttelu (ks. alaluku 5.2), rinnakkaisteksteissä esiintyvä puhuttelu ei noudata yksittäisen tekstin tasolla alaluvussa 5.1 esiteltyä koko rinnakkaistekstiaineiston puhuttelumuotojen jakaumaa, jonka mukaan yleisin puhuttelumuoto koko rinnakkaistekstiaineistossa on possessiivisuffiksi, toiseksi yleisin *sinä*-pronomini, kolmanneksi yleisin toisen persoonan verbinmuoto ja harvinaisin imperatiivi. Jotta puhuttelumuotojen esiintymistä yksittäisissä rinnakkaisteksteissä olisi helpompi tarkastella, alle taulukkoon 12 on laskettu kunkin puhuttelumuotoluokan esiintymien lukumäärä tuhatta sanaa kohti kussakin rinnakkaistekstissä:

	R1	R2	R3	R4	R5	R6	R7	R8	R9
2. pers. verbi / 1 000 sanaa	5	0	1	0	10	16	2	0	23
<i>sinä</i> -pronomini / 1 000 sanaa	11	2	0	0	13	38	7	0	23
Poss.suff. / 1 000 sanaa	3	2	0	3	25	48	5	0	82
Imperatiivi / 1 000 sanaa	0	8	0	0	0	0	0	0	1
Yhteensä	19	12	1	3	48	102	14	0	129

Taulukko 12. Puhuttelumuotoluokkien esiintymät tuhatta sanaa kohti kussakin rinnakkaistekstissä. Luvut on pyöristetty lähimpään kokonaiseen lukuun.

Koko rinnakkaistekstiaineiston jakaumaa noudattavat yksittäisen tekstin tasolla ainoastaan rinnakkaistekstit R5 ja R6. Kaikissa muissa teksteissä puhuttelumuotojen esiintyminen poikkeaa yhden tai useamman puhuttelumuotoluokan osalta koko aineiston jakaumasta. Yksikään rinnakkaisteksti ei kuitenkaan noudata myöskään käänösaineiston jakaumaa. Mielenkiintoista on myös huomata, että ainoastaan yhdessä rinnakkaistekstissä (R9) esiintyy jokaisen puhuttelumuotoluokan edustajia. Teksteistä yhdessä (R8) taas ei esiinny suoraa

puhuttelua lainkaan. Nämä ja aiemmin tässä luvussa kuvatut havainnot tukevat tutkimuksen hypoteesia (ks. luku 1) siinä mielessä, että rinnakkaisteksteissä esiintyy yleisellä tasolla suoraa puhuttelua huomattavasti vähemmän kuin käännöksissä. Analyysin perusteella ei kuitenkaan vaikuta siltä, että puhuttelua olisi ainoastaan vähemmän, vaan puhuttelumuodot esiintyvät rinnakkaisteksteissä eri suhteissa ja hajontaa eri tekstien välillä on huomattavasti, kun niitä verrataan käännösaineistoon, jossa kaikki tekstit noudattavat keskenään samaa jakaumaa.

Seuraavaksi rinnakkaisteksteissä esiintynyttä puhuttelua tarkastellaan vielä aitojen esimerkkien avulla, ja alle esimerkkeihin 8–11 on poimittu puhuttelua sisältäviä katkelmia rinnakkaistekstiaineistosta. Kussakin esimerkissä keskitytään tarkastelemaan yhtä puhuttelumuotoluokkaa kerrallaan, alaluvussa 4.3 esitellyn luokittelun osoittamassa järjestyksessä. Kuten aiemmissakin esimerkeissä, kulloinkin tarkasteltavan puhuttelumuodon sisältämä sana on kursivoitu ja mahdollinen puhuttelua ilmaiseva liite tai pääte alleviivattu. Mikäli katkelmassa esiintyy useamman puhuttelumuotoluokan edustajia, vain kulloinkin tarkasteltava puhuttelumuoto on kursivoitu ja muut jätetty merkitsemättä.

Ensimmäiseksi esimerkissä 8 alla tarkastellaan toisen persoonan verbinmuotoa:

- (8) a. Käyttämällä verkkosivustoamme *hyväksyt* nämä käytännöt. (R6)
- b. Mikäli *et* halua tai voi käyttää portaalin palveluita, *voit* myös lähettää meille henkilökohtaisesti allekirjoittamasi pyynnön kirjeitse yllä mainittuun osoitteeseen. (R9)

Nämä virkkeet ovat esimerkkejä siitä, millaisissa tilanteissa toisen persoonan verbinmuotoja esiintyi rinnakkaistekstiaineistossa. Esimerkkivirkkeet liittyvät ehtojen hyväksymiseen (virke 8 a.) ja lukijan oikeuksien esittelyyn (virke 8 b.). Toisen persoonan verbinmuotoa käytettiin rinnakkaisteksteissä muun muassa yllä olevan tyypisissä tilanteissa, joissa lukijalle kerrottiin tämän oikeuksista tai jostakin tietosuojaselosteeseen yleisesti liittyvästä asiasta (kuten ehtojen hyväksymisestä, yhteyden ottamisesta, tietosuojaselosteen lukemisesta). Toisen persoonan verbinmuotoa käytettiin harvemmin kohdissa, joissa kerrottiin, mitä tietoja lukijasta kerätään (vrt. toisen persoonan verbin esiintyminen käännöksissä, esimerkki 5, alaluku 5.2).

Sinä-pronominia esiintyi rinnakkaistekstiaineistossa toiseksi eniten (ks. taulukot 3 ja 4). Alle on poimittu rinnakkaisteksteistä kaksi persoonapronominilla puhuttelua sisältävää esimerkkiä:

- (9) a. Jos huomaat, että käsittelyssä on puutteita tai se on lainvastaista, *sinulla* on oikeus tehdä asiasta valitus tietosuojaviranomaiselle. (R5)
- b. Keräämme tietojasi myös erilaisten asiakastutkimusten ja -kyselyiden yhteydessä, jotta voisimme vastata *sinulle* ja kehittää palveluitamme. (R9)

Persoonapronomineja käytettiin rinnakkaisteksteissä samantyyppisissä tilanteissa kuin käännöksissäkin (vrt. esimerkki 6, alaluku 5.2) – niillä ilmaistiin lukijan oikeuksia ja kerrottiin, mitä lukijan henkilötietoja kerätään ja mihin tarkoituksiin. Vaikka *sinä*-pronomini oli rinnakkaisteksteissä toiseksi yleisin puhuttelun muoto, sitä esiintyi vain kuudessa tekstissä yhdeksästä, eikä sitä esiintynyt rinnakkaisteksteissä yhtä usein kuin käännöksissä (rinnakkaisteksteissä 7 / 1 000 sanaa, käännöksissä 18 / 1 000 sanaa). Käännöksissä *sinä*-pronomini esiintyi siis yli kaksi kertaa useammin kuin rinnakkaisteksteissä. On kuitenkin mielenkiintoista, että *sinä*-pronomini esiintyy rinnakkaisteksteissä useammin kuin toisen persoonan verbinmuoto, sillä suomen kielelle on ominaisempaa jättää nimenomaan persoonapronomini pois ja osoittaa puhuttelu esimerkiksi pelkällä verbin taivutuksella (ks. alaluku 2.2). Toisaalta taas yllä olevien esimerkkilauseiden kaltaisissa yhteyksissä *sinä*-pronominiin ei liity persoonamuotoista verbiä tai substantiivia, jolloin pronominia ei edes voisi jättää pois. Täytyy myös huomata, että tutkimusaineistossani yksi yksittäinen teksti sisältää lähes puolet rinnakkaistekstiaineiston *sinä*-pronominin esiintymistä (ks. taulukko 9). Kyseinen teksti R9 on muiltakin osin hyvin poikkeava muihin rinnakkaisteksteihin verrattuna, mikä herättää kysymyksiä tekstin alkuperästä ja luotettavuudesta tämän tutkimuksen kannalta. Asiaa pohditaan vielä lyhyesti alaluvussa 5.4.

Esimerkissä 10 alla tarkastellaan possessiivisuffiksia rinnakkaistekstiaineistosta poimitun aidon tekstikatkelman avulla. Possessiivisuffiksi oli koko rinnakkaistekstiaineistossa yleisin suoran puhuttelun muoto, mutta yksittäisen tekstin tasolla se oli yleisin ainoastaan kolmessa yhdeksästä tekstistä.

- (10) Voimme käsitellä *henkilötietojasi* oikeutetun edun perusteella osana liiketoimintaamme harkittuamme, ettei käsittely loukkaa *henkilötietojesi* suojaa. Näissä tilanteissa sinulla on oikeus vastustaa *henkilötietojesi* käsittelyä henkilökohtaisiin *syihisi* vedoten. (R9)

Possessiivisuffiksi esiintyi rinnakkaisteksteissä usein esimerkiksi sanaan *henkilötiedot* liitettynä (ks. tarkemmin taulukko 13 jäljempänä). Possessiivisuffikseja käyttämällä korostettiin lukijalle siis sitä, että kyseessä ovat juuri hänen henkilötietonsa ja oikeutensa. Possessiivisuffiksi liittyi rinnakkaisteksteissä useammin substantiiviin (esimerkiksi *oikeutesi*)

kuin verbin infiniittimuotoon (esimerkiksi *ostaessasi*), kun taas käänöksissä verbiinkin liitettyjä possessiivisuffikseja oli runsaasti (ks. taulukko 9, alaluku 5.2).

Viimeisenä esimerkissä 11 tarkastellaan imperatiivia, jota esiintyi suoran puhuttelun muodoista rinnakkaisteksteissä vähiten. Sitä käytettiin vain kahdessa yhdeksästä tarkastellusta tekstistä:

- (11) a. *Lue* lisää tiedon käsittelystä tuotekehitystä varten täältä. (R2)
- b. Mikäli et halua tai voi käyttää portaalin palveluita, voit myös lähettää meille henkilökohtaisesti allekirjoittamasi pyynnön kirjeitse yllä mainittuun osoitteeseen. *Merkitse* kuoreen ”tietosuoja”. (R9)

Rinnakkaistekstien kuudesta esiintymästä viidessä imperatiivina oli sana *lue*, ja kerran esiintyi sana *merkitse*. Imperatiivilla siis lähinnä ohjeistettiin lukijaa lisätoimiin, mikäli lukija tämän katsoisi tarpeelliseksi.

Jotta aineistosta ja siinä esiintyvistä suorasta puhuttelusta saataisiin mahdollisimman konkreettinen käsitys, taulukkoon 13 alla on koottu jokaisen puhuttelumuotoluokan osalta kolme yleisintä sanaa, johon puhuttelumuoto liittyy (*sinä*-pronominin osalta kolme yleisintä taivutusmuotoa):

	2. persoonan verbinmuoto (esiintymiä)	<i>sinä</i> -pronomini (esiintymiä)	Possessiivisuffiksi (esiintymiä)	Imperatiivi (esiintymiä)
Yleisin	<i>voida</i> (10)	<i>sinulla</i> (14)	<i>henkilötieto</i> (14)	<i>lukea</i> (5)
2. yleisin	<i>huomata</i> (3)	<i>sinulle</i> (10)	<i>tieto</i> (9)	<i>merkitä</i> (1)
3. yleisin	<i>olla</i> (2)	<i>sinua</i> (5)	<i>antaa</i> (4)	

Taulukko 13. Kolme yleisintä sanaa, johon puhuttelumuoto liittyy, kunkin puhuttelumuotoluokan osalta rinnakkaisteksteissä.³

Taulukosta nähdään rinnakkaistekstiaineiston yleisimmät puhuttelua sisältävät sanamuodot ja sanat, joihin puhuttelumuodot liittyvät. Verbeistä selvästi yleisin oli *voida* (esimerkiksi ”*voit* peruuttaa suostumuksesi” tai ”*voit* ottaa yhteyttä”), toiseksi yleisin oli *huomata* (esimerkiksi ”jos *huomaat*, että käsittelyssä on puutteita”) ja kolmanneksi yleisin *olla* (esimerkiksi ”jos *olet* vastustanut henkilötietojesi käsittelyä”). *Sinä*-pronomini taas esiintyi useimmiten adessiivimuodossa *sinulla* (aina merkityksessä ”*sinulla* on oikeus”), allatiivimuodossa *sinulle*

³ Imperatiivin osalta annettu vain kaksi yleisintä sanaa, sillä aineistossa ei esiintynyt muita sanoja imperatiivissa.

(esimerkiksi ”ilmoitamme *sinulle*” tai ”tietoja *sinulle* kohdennetusta markkinoinnista”) tai partitiivimuodossa *sinua* (esimerkiksi ”*sinua* koskevat henkilötiedot”). Possessiivisuffiksi taas liittyi useimmiten substantiiviin *henkilötieto* (esimerkiksi ”säilytämme *henkilötietojasi*”) ja substantiiviin *tieto* (esimerkiksi ”käsittelemme *tietojasi*” tai ”keräämme *tietojasi*”). Possessiivisuffikseja ei esiintynyt verbien yhteydessä niin usein kuin substantiivien yhteydessä, mutta esimerkiksi *antaa*-verbin infiniittimuotojen yhteydessä niitä esiintyi välillä (esimerkiksi ”*antamasi* suostumus”). Imperatiivi esiintyi koko aineistossa vain kuusi kertaa, muodoissa *lue* ja *merkitse* (esimerkiksi ”*lue* tarkemmin” tai ”*merkitse* kuoreen ’tietosuojaa’”), minkä vuoksi taulukkoon ei ole merkitty kolmanneksi yleisintä imperatiivimuotoa.

Vaikka tässä tutkimuksessa on keskitytty nimenomaan suoraan puhutteluun tietosuojaselosteissa, on hyvä huomata, että teksteissä viitataan lukijaan myös epäsuorasti. Jotta ei syntyisi vaikutelmaa, että tutkimissani tietosuojateksteissä viitataan lukijaan vain suoralla puhuttelulla, alle esimerkkiin 12 on poimittu joitakin esimerkkejä lukijaa tarkoittaviksi tulkittavista epäsuorista viittauksista kolmanteen persoonaan. Käännösaineistossa vastaavanlaisia epäsuoria viittauksia oli vain muutama yksittäinen rajatapaus, joten ilmiötä ei ole mielekästä tarkastella käännösaineiston osalta. Rinnakkaisteksteissä epäsuoria viittauksia oli runsaasti, ja alla olevissa esimerkkilauseissa epäsuorat viittaukset lukijaan on kursivoitu:

- (12) a. *Asiakkaalla* on oikeus saada rekisterinpitäjältä vahvistus siitä, käsitelläänkö *häntä* koskevia henkilötietoja. *Asiakkaalla* on lisäksi oikeus saada jäljennös *häntä* koskevista henkilötiedoista sekä tietosuojasetuksen mukaiset tiedot henkilötietojen käsittelystä. (R8)
- b. Jos verkkopalveluiden *käyttäjä* ei hyväksy tätä tietosuojakäytäntöä taikka [– –] verkkopalveluiden käyttöehtoja, verkkopalveluita ei tule käyttää. (R1)
- c. *Rekisteröidyn* oikeuksien toteuttamista koskevat tarkemmat ohjeet löytyvät osoitteesta [– –]. *Rekisteröity* voi pyytää jäljennöksen *henkilötiedoistaan* maksutta kerran vuodessa. (R3)

Muun muassa yllä olevissa esimerkkivirkkeissä esiintyviä sanoja *asiakas*, *käyttäjä* ja *rekisteröity* käytettiin rinnakkaisteksteissä viittaamaan lukijaan epäsuorasti. Persoonaviittausten välttäminen on suomen kielelle hyvin ominaista (Hakulinen 2001 [1987]: 209), joten tällainen suoran puhuttelun korvaaminen epäsuorilla viittauksilla sopii suomen kieleen hyvin, eikä herätä lukijassa ihmetystä. Toisaalta Isosävin ja Lappalaisen (2015: 12) mukaan nimenomaan kolmannen persoonan puhuttelusanojen käyttö ei ole ainakaan suomenkielisessä puheessa kovin yleistä. Tutkimukseni aineisto on siis tässä mielessä poikkeuksellista. On kenties mahdollista, että kolmannen persoonan puhuttelua käytetään

puheessa varsin harvoin, mutta teksteissä useammin. Ero käännösten ja rinnakkaistekstien välillä on tältä osin merkittävä, sillä kuten mainittua, epäsuoraa puhuttelua ei käännöksissä esiinny käytännössä ollenkaan.

5.4 Päätelmät

Edellä alaluvuissa 5.1–5.3 esiteltiin analyysin tulokset niin yleisellä kuin yksittäistenkin tekstien tasolla. Yleisellä tasolla tarkasteltuna suoraa puhuttelua esiintyi käännöksissä huomattavasti rinnakkaistekstejä enemmän, minkä voi ainakin osittain olettaa olevan lähtökielen vaikutusta eli *interferenssiä*. Toisaalta suoran puhuttelun runsasta määrää voisi käänteisesti selittää myös *uniikkiaines*-käsite. Tällä tarkoitan sitä, että kun uniikkiainesten ajatellaan yleensä olevan jotakin sellaista, mihin lähtötekstissä ei ole kimmoketta, tutkimassani aineistossa ilmiö olisikin käänteinen ja lähtöteksti antaisi kimmokkeen käyttää suoraa puhuttelua useammin kuin alun perin suomeksi kirjoitettaessa. Lähtötekstin vaikutuksesta ei kuitenkaan voida päästä absoluuttiseen varmuuteen, sillä tutkimusaineistoni ei sisältänyt alkutekstejä. Aiemmassa tutkimuksessa on kuitenkin myös tehty samansuuntaisia havaintoja käännetyin suomen kielestä (esimerkiksi Mauranen & Tiittula 2005), joten todennäköistä on, että runsas puhuttelu on ainakin jossakin määrin lähtökielen vaikutusta.

Tuloksia tarkasteltaessa on lisäksi pidettävä mielessä, ettei lähtötekstien kielestä tai rinnakkaistekstien alkuperästä voida päästä tämän tutkimuksen puitteissa täyteen varmuuteen. Vaikka luultavaa on, että käännösaineiston tekstit ovat käännöksiä nimenomaan englannista ja rinnakkaistekstiaineiston tekstit on kirjoitettu alun perin suomeksi, tästä ei ole täyttä varmuutta, sillä käytettävissä ei ollut tietoa tekstien laatijoista tai kääntäjistä. Näin ollen myöskään oletuksia esimerkiksi yksittäisen kääntäjän käännösratkaisuista ei voi analyysin ja aineiston perusteella tehdä. Tutkimuksen tarkoituksena oli kuitenkin nimenomaan sen selvittäminen, millaista lukijan suora puhuttelu on käännöksissä verrattuna rinnakkaisteksteihin, ja tähän kysymykseen saatiin tutkimuksen avulla kaksi pääasiallista vastausta: suora puhuttelu eroaa käännösten ja rinnakkaistekstien välillä (1) esiintymistiheyden osalta ja (2) käytettyjen muotojen osalta.

Analyysin perusteella käännösaineisto näyttää puhuttelun osalta melko homogeeniselta, kun taas rinnakkaistekstien välillä on enemmän keskinäistä hajontaa. Esimerkiksi rinnakkaisteksti R9 oli puhuttelultaan huomattavasti lähempänä käännöksiä kuin muita rinnakkaistekstejä. Tämän tutkimuksen puitteissa ei ollut mahdollista selvittää, mistä tällainen hajonta johtuu,

mutta mahdollisia syitä voisivat olla esimerkiksi se, että johonkin tekstiin on otettu mallia käännöksestä tai vieraskielisestä tekstistä tai jopa se, että jokin rinnakkaisteksti tai -tekstin osa onkin itse asiassa käännös. Toisaalta kyse voi myös olla yksittäisen kirjoittajan henkilökohtaisista mieltymyksistä. Vastaavasti taas käännösaineiston homogeenisyyteen on saattanut vaikuttaa englannin kielen pronominin *you* yleisyys ja se, ettei sille juurikaan ole nykyenglannissa vaihtoehtoisia ilmaisuja, jolloin sama homogeenisyys välittyy käännöksiin. Tekstien alkuperään liittyvät seikat voivat tietenkin vaikuttaa tutkimuksen tulosten luotettavuuteen ja yleistettävyyteen, mutta tätä mahdollisuutta pyrittiin minimoimaan keräämällä tarpeeksi suuri tutkimusaineisto ja valitsemalla aineiston tekstit eri lähteistä.

Tämä tutkimus antaa osviittaa siitä, miten lukijaa puhutellaan alun perin suomeksi kirjoitetuissa tietosuojaselosteissa ja miten taas englannista suomeksi käännettyissä tietosuojaselosteissa. Tutkimusaineistossani ero käännösten ja rinnakkaistekstien välillä oli huomattava, ja yleisellä tasolla tästä voidaan päätellä, että englannista suomeen käännettyissä teksteissä suora puhuttelu on jonkin verran yleisempää kuin suomeksi alun perin kirjoitetuissa teksteissä, mutta laajoja yleistyksiä ei tämänlaajuisen ja yhteen tekstilajiin keskittyvän aineiston pohjalta ole mahdollista tehdä. Tulokset kuitenkin kertovat jotakin suomen kielen puhuttelunormeista ja tietosuojaselosteiden kääntämisestä sekä mahdollisesti käytännön kääntämisestä hieman yleisemmälläkin tasolla.

Tutkimuksen tulokset viittaavat mielestäni siihen, mitä Lappalainenkin (2015: 100; ks. myös tämän tutkielman alaluku 2.2) totesi tutkimuksessaan suomen kielen puhuttelukäytännöistä: suomen kielessä ei ole selkeitä puhuttelunormeja. Tästä kertoo yhtäältä se, että tutkimusaineistoni rinnakkaisteksteissä oli puhuttelun osalta merkittävää keskinäistä vaihtelua (suoran puhuttelun esiintymien osuus yksittäisen tekstin sanamäärästä vaihteli välillä 0–12,9 %), ja toisaalta se, että rinnakkaistekstien ja käännösten välillä oli puhuttelun kokonaismäärän osalta merkittävä ero (rinnakkaisteksteissä suoran puhuttelun esiintymien osuus aineiston sanamäärästä oli 3,1 %, käännöksissä 11,6 %). Jos puhuttelunormit olisivat selkeät, vaihtelua ei luultavasti olisi kummassakaan suhteessa näin paljoa, kun kirjoittajilla ja kääntäjillä olisi selkeä tieto siitä, millaisia puhutteluilmauksia kuuluu milloinkin käyttää.

Tulosten perusteella vaikuttaa myös siltä, ettei tietosuojaselosteelle tekstilajina ole vielä muodostunut yhtenäisiä konventioita. Tätäkin päätelmää tukevat yllä kuvatut vaihtelut suoran puhuttelun määrästä niin rinnakkaistekstiaineiston sisällä kuin rinnakkaistekstien ja käännösten välilläkin. Rinnakkaistekstien keskinäinen vaihtelu kertoo siitä, että tietosuojaselosteelle ei

tekstilajina ole vielä muodostunut puhuttelukonventioita. Eron rinnakkaistekstien ja käännösten välillä voisi ajatella kielivän samasta asiasta, sillä käännösten voisi ajatella mahdollisesti noudattavan lähtötekstin rakenteita kohdekielen konventioiden puuttuessa. Toisaalta taas on mahdollista, että lähtötekstin rakenteiden mahdollinen noudattaminen kertoisi siitä, että käännökset on tehty kiireessä ja pienellä budjetilla. Tämä on kuitenkin välillä valitettava totuus käytännön käännösmaailmassa, joten on hyvä muistaa, ettei tämänkään tutkimuksen käännösaineiston käännöksiä ole tehty laboratorio-olosuhteissa loputtomien resurssien ja ilman aikarajoituksia. Tietosuojavaluuttu (2019) on myös antanut ohjeen siitä, että tietosuojaselosteiden eri kieliversioiden tulee vastata toisiaan (ks. alaluku 4.1.3), mikä saattaisi myös olla omiaan siirtämään lähtötekstin rakenteita kohdetekstiin. Lisäksi on tietenkin myös mahdollista, että runsas sinuttelu kielii Larjavaaran (1999) kuvaamasta yhteiskunnallisesta ja asenteellisesta muutoksesta (ks. alaluku 2.2). Tämän tulkinnan kanssa olisin kuitenkin hyvin varovainen, sillä yhteiskunnallisen ja asenteellisen muutoksen voisi olettaa heijastuvan myös alun perin suomeksi kirjoitettuihin teksteihin.

Mahdollisia tulkintoja tulosten merkityksestä on siis monia, joista ainakin osan todentaminen vaatisi jatkotutkimuksia. Seuraavaksi luvussa 6 tehdään vielä lyhyt yhteenveto tutkimuksesta ja luodaan katsaus mahdollisiin jatkotutkimusaiheisiin.

6 Lopuksi

Tässä pro gradu -tutkielmassa on tarkasteltu lukijan suoraa puhuttelua eli tapoja kohdistaa teksti suoraan lukijalle suomenkielisissä *tietosuojaselosteissa*. Tutkimusaineisto koostui kuudesta englannista suomeen käännetyistä tietosuojaselosteesta ja yhdeksästä alun perin suomeksi kirjoitetusta tietosuojaselosteesta eli *rinnakkaistekstistä*. Tietosuojaselosteella tarkoitettiin tässä tutkimuksessa tekstiä, jossa henkilötietoja keräävä ja käsittelevä taho kuvaa kerättyjen tietojen kohteelle Euroopan unionin yleisen tietosuoja-asetuksen (EU) 2016/679 mukaiset tiedot muun muassa tietojen käsittelystä ja keräämisestä sekä tietojen kohteen oikeuksista.

Tutkimusasetelma oli vertaileva: tutkimuksessa vertailtiin lukijan suoran puhuttelun tapoja käännöksissä ja rinnakkaisteksteissä. Tavoite oli selvittää tietosuojaselosteissa esiintyvää suoraa puhuttelua hakemalla aineiston 15 tekstistä vastausta kysymykseen millä tavoin, jos millään, lukijan suora puhuttelu eroaa englannista suomeen käännetyissä tietosuojaselosteissa verrattuna alun perin suomeksi kirjoitettuihin tietosuojaselosteisiin. Tutkimuskysymykseen etsittiin vastausta poimimalla ensin aineistosta kaikki suoran puhuttelun muodot ja luokittelemalla ne Eveliina Tolvasen (2016) käyttämän luokittelun mukaisesti. Käytin Tolvasen luokittelusta hieman muokattua versiota, jonka luokat olivat seuraavat: (1) toisen persoonan verbinmuoto, (2) *sinä*-pronomini, (3) possessiivisuffiksi ja (4) imperatiivi.

Tutkimuksen hypoteesina oli, että englanninkielisten lähtötekstien vaikutuksen voisi havaita suomenkielisissä käännöksissä esimerkiksi siinä, että suoraa puhuttelua on käännöksissä rinnakkaistekstejä enemmän. Hypoteesini osoittautui pitkälti oikeaksi, sillä käännösaineistossa esiintyi suoraa puhuttelua merkittävästi rinnakkaistekstejä enemmän. Käännösaineiston kokonaissanamäärästä noin 11,6 prosenttia oli suoraa puhuttelua sisältäviä muotoja, kun taas rinnakkaistekstiaineiston osalta vastaava luku oli 3,1. Eroja oli myöskin siinä, mitä muotoja suosittiin. Oletettavaa on, että tulos selittyy ainakin osittain lähtökielen vaikutuksella. Tästä ei kuitenkaan voida saada täyttä varmuutta, sillä tutkimusaineisto ei sisältänyt lähtötekstejä. Tutkimuksen tulokset tukevat aiempaa tutkimusta käännösten ja etenkin käännetyin suomen kielestä yleisellä tasolla: muun muassa Mauranen ja Tiittula (esimerkiksi 2005) ovat aiemmin todenneet, että käännetyssä suomessa esiintyy enemmän persoonaviittauksia kuin alkuperäissuomessa.

Tutkimusaiheeni tarjoaa myös monenlaisia jatkotutkimusmahdollisuuksia. Tulevaisuudessa olisi mielenkiintoista perehtyä esimerkiksi puhuttelumuotojen kollokaatteihin eli siihen, mitkä

sanat esiintyvät usein puhuttelumuotojen läheisyydessä. Lisäksi olisi antoisaa tutkia tietosuojaselosteissa esiintyvää epäsuoraa puhuttelua tarkemmin. Tulosten yleistettävyyden parantamiseksi suoraa puhuttelua tulisi myös tutkia laajemmasta aineistosta ja useammista tekstilajeista ja interferenssihypoteesin varmistamiseksi mukaan olisi otettava myös lähtötekstit.

Lähteet

Tutkimusaineisto

Käännökset

- K1** = Apple 2018. *Tietosuojakäytäntö*. Saatavissa: <https://www.apple.com/legal/privacy/fi/>. [Luettu 12.2.2019.]
- K2** = Netflix 2018. *Tietosuojalausunto*. Saatavissa: <https://help.netflix.com/legal/privacy>. [Luettu 12.2.2019.]
- K3** = Microsoft 2018. *Tietosuojalauseke*. Saatavissa: <https://privacy.microsoft.com/fi-fi/privacystatement>. [Luettu 12.2.2019.]
- K4** = Facebook 2018. *Tietokäytäntö*. Saatavissa: <https://www.facebook.com/privacy/explanation>. [Luettu 12.2.2019.]
- K5** = McDonald's 2018. *McDonald'sin tietosuojaseloste*. Saatavissa: <http://www.mcdonalds.fi/fi/rekisteriseloste.html>. [Luettu 12.2.2019.]
- K6** = Electronic Arts 2018. *Tietosuoja- ja evästekäytäntö*. Saatavissa: <http://tos.ea.com/legalapp/WEBPRIVACY/FI/fi/PC/>. [Luettu 19.2.2019.]

Rinnakkaistekstit

- R1** = Atria 2018. *Tietosuojakäytäntö*. Saatavissa: <https://www.atria.fi/tietosuojakaytanto/>. [Luettu 12.2.2019.]
- R2** = Sanoma 2018. *Tietosuojalauseke*. Saatavissa: <https://sanoma.fi/tietoa-meista/tietosuoja/tietosuojalauseke/#osa-1>. [Luettu 12.2.2019.]
- R3** = HSL 2018. *HSL:n tietosuoja*. Saatavissa: <https://www.hsl.fi/tietosuoja>. [Luettu 12.2.2019.]
- R4** = Alko 2018. *Tietosuojaseloste*. Saatavissa: <https://www.alko.fi/verkkokaupan-tietosuojaseloste>. [Luettu 12.2.2019.]
- R5** = Verkkokauppa.com 2018. *Tietosuojaseloste*. Saatavissa: <https://www.verkkokauppa.com/fi/ohjeet/tietosuojaseloste>. [Luettu 12.2.2019.]
- R6** = Olvi 2018. *Tietosuoja- ja evästekäytännöt*. Saatavissa: <https://www.olvi.fi/yhtio/vastuullisuus/yleista/tietosuoja-ja-evastekaytannot/>. [Luettu 12.2.2019.]
- R7** = Evidensia 2018. *Asiakasrekisterin rekisteri- ja tietosuojaseloste*. Saatavissa: <https://www.evidensia.fi/asiakasrekisterin-rekisteri-ja-tietosuojaseloste/>. [Luettu 15.2.2019.]
- R8** = Musti ja Mirri 2017. *Tietosuojaseloste*. Saatavissa: <https://www.mustijamirri.fi/rekisteriseloste#5>. [Luettu 12.2.2019.]
- R9** = K-Plussa 2018. *Tietosuojaseloste*. Saatavissa: <https://www.plussa.com/Mika-on-K-Plussa/Sopimusehdot/Rekisteriselosteen/>. [Luettu 15.2.2019.]

Tutkimuskirjallisuus

- Baker, Mona 1993. Corpus Linguistics and Translation Studies – Implications and Applications. Teoksessa: Mona Baker, Gill Francis & Elena Tognini-Bonelli (toim.), *Text and Technology – In Honour of John Sinclair*. Amsterdam/Philadelphia: John Benjamins Publishing Company. 233–250. [Luettu e-kirjana 30.4.2019.]
- Braun, Friederike 1988. *Terms of Address – Problems of patterns and usage in various languages and cultures*. Berlin: Mouton de Gruyter.
- Brown, Penelope & Stephen C. Levinson 1987. *Politeness. Some Universals in Language Use*. Cambridge: Cambridge University Press. [Julkaistu alun perin vuonna 1978 teoksessa Esther N. Goody (toim.), *Questions and Politeness*.]
- Brown, Roger & Albert Gilman 1960. The Pronouns of Power and Solidarity. Teoksessa: Thomas A. Sebeok (toim.), *Style in Language*. Toinen painos. Cambridge, Massachusetts: The M.I.T Press. 253–276.
- Clyne, Michael & Catrin Norrby & Jane Warren 2009. *Language and Human Relations – Styles of Address in Contemporary Language*. Cambridge: Cambridge University Press.
- Euroopan unionin yleinen tietosuoja-asetus (EU) 2106/679. Saatavissa: <https://eur-lex.europa.eu/legal-content/FI/TXT/?uri=CELEX%3A32016R0679>. [Luettu 4.5.2019.]
- Hakulinen, Auli 2001 [1987]. Persoonaviittauksen välttäminen suomessa. [Englanninkielinen alkuperäisartikkeli *Avoiding Personal Reference in Finnish*, suomennos Lea Laitinen.] Teoksessa: Lea Laitinen, Pirkko Nuolijärvi, Marja-Leena Sorjonen & Maria Vilks (toim.), *Lukemisto. Kirjoituksia kolmelta vuosikymmeneltä*. Suomalaisen Kirjallisuuden Seuran toimituksia 816. Helsinki: Suomalaisen Kirjallisuuden Seura. 209–208.
- Hallintolaki 6.6.2003/434. Saatavissa: <http://www.finlex.fi/fi/laki/ajantasa/2003/20030434?search%5Btype%5D=pika&search%5Bpika%5D=hallintolaki>. [Luettu 4.5.2019.]
- Isosävi, Johanna & Hanna Lappalainen 2015: Johdanto. Teoksessa: Johanna Isosävi & Hanna Lappalainen (toim.), *Saako sinutella vai täytyykö teititellä? Tutkimuksia eurooppalaisten kielten puhuttelukäytännöistä*. Tietolipas 246. Helsinki: Suomalaisen Kirjallisuuden Seura. 9–33.
- Jespersen, Otto 1952. *A Modern English Grammar on Historical Principles*. 5, Syntax. Fourth Volume. Kööpenhamina: Ejnar Munksgaard.
- Jucker, Andreas H. & Irma Taavitsainen 2013. *English Historical Pragmatics*. Edinburgh: Edinburgh University Press. [Luettu e-kirjana 16.4.2019.]
- Lappalainen, Hanna 2015. *Sinä vai te vai sekä että?* Teoksessa: Johanna Isosävi & Hanna Lappalainen (toim.), *Saako sinutella vai täytyykö teititellä? Tutkimuksia eurooppalaisten kielten puhuttelukäytännöistä*. Tietolipas 246. Helsinki: Suomalaisen Kirjallisuuden Seura. 72–104.
- Larjavaara, Matti 1999. Kieli, kohteliaisuus ja puhuttelu. *Kielikello* 2/1999. Saatavissa: <https://www.kielikello.fi/-/kieli-kohteliaisuus-ja-puhuttelu>. [Luettu 27.4.2019.]
- Mauranen Anna 2004. Corpora, universals and interference. Teoksessa: Anna Mauranen & Pekka Kujamäki (toim.), *Translation Universals – Do They Exist?* Amsterdam/Philadelphia: John Benjamins Publishing Company. 65–82. [Luettu e-kirjana 27.4.2019.]

- Mauranen, Anna & Pekka Kujamäki 2004. Introduction. Teoksessa: Anna Mauranen & Pekka Kujamäki, *Translation Universals – Do They Exist?* Amsterdam/Philadelphia: John Benjamins Publishing Company. [Luettu e-kirjana 16.4.2019.]
- Mauranen, Anna & Liisa Tiittula 2005. *Minä käännössuomessa ja supisuomessa*. Teoksessa: Anna Mauranen & Jarmo H. Jantunen (toim.), *Käännössuomeksi. Tutkimuksia suomennosten kielestä*. Tampere Studies in Language, Translation and Culture Sarja A Vol. 1. Tampere: Tampere University Press. 35–70.
- Nevala, Minna 2015. Elisabetista Elisabetiin – Englannin puhuttelukäytänteiden sosiaalisesta ulottuvuudesta. Teoksessa: Johanna Isosävi & Hanna Lappalainen (toim.), *Saako sinutella vai täytyykö teititellä? Tutkimuksia eurooppalaisten kielten puhuttelukäytännöistä*. Tietolipas 246. Helsinki: Suomalaisen Kirjallisuuden Seura. 262–286.
- Oikeusasiamies 2014. *Asiakkaan sinuttelu Kelan vakiokirjeessä ei ole vastoin lakia*. Saatavissa: <https://www.oikeusasiamies.fi/fi/ratkaisut/-/eoar/4505/2013>. [Luettu 21.4.2019.]
- Pym, Anthony 2011. Translation research terms: A tentative glossary for moments of perplexity and dispute. *Translation Research Projects* 3. 75–99. Saatavissa: https://www.researchgate.net/publication/283363357_Translation_research_terms_A_tentative_glossary_for_moments_of_perplexity_and_dispute. [Luettu 5.5.2019.]
- Tietosuojaavaltuutettu 2019. *Kerro käsittelystä rekisteröidylle*. Saatavissa: <https://tietosuoja.fi/rekisteroidyn-informointi>. [Luettu 21.2.2019.]
- Tirkkonen-Condit, Sonja 2002. Translationese – a myth or an empirical fact? A study into the linguistic identifiability of translated language. *Target* 14 (2). 207–220. Saatavissa: <http://docserver.ingentaconnect.com/deliver/connect/jbp/09241884/v14n2/s1.pdf?expires=1555151920&id=0000&titleid=4151&checksum=D0376ED9CFA1B9C05F8A01201A85FF9E>. [Luettu 27.4.2019.]
- Tirkkonen-Condit, Sonja 2004. Unique items – over- or under-represented in translated language? Teoksessa: Anna Mauranen & Pekka Kujamäki, *Translation Universals – Do They Exist?* Amsterdam/Philadelphia: John Benjamins Publishing Company. [Luettu e-kirjana 16.4.2019.]
- Tirkkonen-Condit, Sonja 2005. Häviävätkö uniikkiaineokset käännössuomesta? Teoksessa: Anna Mauranen & Jarmo H. Jantunen (toim.), *Käännössuomeksi. Tutkimuksia suomennosten kielestä*. Tampere Studies in Language, Translation and Culture Sarja A Vol. 1. Tampere: Tampere University Press. 123–137.
- Tolvanen, Eveliina 2016. Lukijan suora puhuttelu suomalaisissa ja ruotsinsuomalaisissa viranomaisteksteissä. *Virittäjä* 1/2016. 5–29. Saatavissa: <https://journal.fi/virittaja/article/view/49565>. [Luettu 21.2.2019.]
- Toury, Gideon 1995. *Descriptive Translation Studies – And Beyond*. Neljäs painos. Amsterdam/Philadelphia: John Benjamins Publishing Company. [Luettu e-kirjana 30.4.2019.]
- VISK = Hakulinen, Auli & Maria Vilkkuna, Riitta Korhonen, Vesa Koivisto, Tarja Riitta Heinonen & Irja Alho 2004. *Iso suomen kielioppi*. Verkkoaversio. Helsinki: SKS. Saatavissa: <http://scripta.kotus.fi/visk>. [Luettu 21.2.2019.]
- Weinreich, Uriel 1979. *Languages in Contact – Findings and Problems*. Haag: Mouton Publishers. [Luettu e-kirjana 16.4.2019.]

ENGLISH SUMMARY

University of Helsinki

Faculty of Arts

Master's Programme in Translation and Interpreting

Lotta Tuure: Direct Forms of Address in Privacy Statements – A Comparative Analysis of Finnish Translations and Originally Produced Texts

Master's Thesis 46 pages, English Summary 7 pages

May 2019

Direct Forms of Address in Privacy Statements

A Comparative Analysis of Finnish Translations and Originally Produced Texts

1 Introduction

The European Union's General Data Protection Regulation, or the GDPR, entered into force on 25 May 2018. As a result, new requirements were set in place for data processors, or *data controllers*, regarding the collection, processing, and storage of personal data. One of these requirements is that information on the processing of personal data must be provided to *data subjects*, i.e. the subjects of the collected data, using clear and intelligible language (Tietosuojavalvutettu [Finnish Data Protection Ombudsman] 2019). Consequently, it is important to study the language of privacy statements, translated and originally produced texts, i.e. non-translations, alike. This thesis will focus on one particular element of language in privacy statements – forms of address. More precisely, this study will consider the differences and similarities in the use of direct forms of address between privacy statements translated from English into Finnish and privacy statements originally written in Finnish.

Forms of address (Finnish: '*puhuttelu*') are used in speech or writing to indicate the recipient of what is being said or written (VISK, määritelmät, s.v. *puhuttelu*). Thus, in the context of privacy statements, forms of address are used to indicate whose data is being collected and whom the rights mentioned in the text concern. Forms of address are not universal but rather vary according to language, which may result in translations deviating from originally produced texts in their use of forms of address as a consequence of, for instance, interference (Isosävi & Lappalainen 2015: 11; Mauranen & Tiittula 2005: 36). This study will focus on finding such deviations in translated privacy statements as compared to originally produced texts. The study

material consists of six translations and nine originally produced texts, and the study method is comparative. The theoretical background of this study consists of literature on forms of address and their potential functions, politeness, and the general features of translated language. The next section provides an overview of the theoretical background of this thesis.

2 Theoretical Background

2.1 Forms of Address

As stated earlier, forms of address are used to indicate the recipient of what is being said or written. Furthermore, forms of address affect and construct the relationships between the interacting parties, and they are even said to mirror the social structure of the surrounding society (Clyne, Norrby & Warren 2009: 1). Forms of address are also closely related to social norms and politeness (Isosävi & Lappalainen 2015: 9, 13).

Different languages have different means of expressing direct address. Finnish, for example, uses verb inflections, possessive suffixes, and personal pronouns for direct address. All of these can be used on their own or in conjunction with each other. Finnish also makes a difference between formal and informal address by means of second-person pronouns; generally speaking, ‘*sinä*’ is used for informal address, and ‘*te*’ for formal address. (Isosävi & Lappalainen 2015: 11.) It is also common for Finnish to avoid using direct forms of address altogether (Hakulinen 2001 [1987]: 209).

English uses direct address a lot more frequently than Finnish (Isosävi & Lappalainen 2015: 11), and there is practically only one way of expressing direct address: the pronoun ‘you’, which is a neutral expression and is used a lot (Clyne, Norrby & Warren 2009: 1–4). Other means, such as titles or a first name, are used to achieve a more formal or informal tone (*ibid.*, 39). English used to differentiate between formal and informal address by means of the pronouns ‘ye’/‘you’ and ‘thou’ but ever since the 18th century, the prevailing form has been ‘you’ (Jucker & Taavitsainen 2013: 76–83).

2.2 General Features of Translated Finnish

As the purpose of this study is to compare translated texts with originally produced texts, it is also relevant to observe some general features of translated language, particularly Finnish. Sonja Tirkkonen-Condit (e.g. 2005) has extensively studied *unique items* in translations,

meaning such items in a target language that are not present in the source language. According to Tirkkonen-Condit (2005: 123), translations tend to contain fewer unique items than originally produced texts. The notion of unique items is relevant in terms of this study because some forms of address, such as using inflected verb forms and possessive suffixes alone to express direct address, could be seen as unique items of Finnish as compared to English. By way of example, this could show in a translation as, for instance, the phrase ‘your personal information’ being translated into Finnish using a combination of a pronoun and a possessive suffix (‘*sinun henkilötietosi*’) even though an originally produced text would probably only use a possessive suffix (‘*henkilötietosi*’).

Another general feature of translated language is *source language interference*, i.e. the influence of the source language on the language of the target text (Mauranen & Kujamäki 2004: 66). Interference could be seen as the opposite of unique items. While unique items are something that is present in the target language but not in the source language, interference brings something from the source language into the target language that would not otherwise be there. It has been observed by Anna Mauranen and Liisa Tiittula (2005: 36), among others, that first-person and second-person pronouns are more prevalent in Finnish translations than in texts originally written in Finnish. Based on this observation, the assumption in this study is that source language interference can be seen in the use of address forms in the translations – as English uses direct address more frequently than Finnish, it can be assumed that direct address is also more frequent in the translations as compared to the texts originally written in Finnish.

3 Material and Method

The study material of this thesis consists of translated and original privacy statements. The GDPR regulates the content of privacy statements, which is why all privacy statements are quite similar in content to a certain extent. Pursuant to the GDPR, a privacy statement must always include information regarding the type of data being collected, the grounds and purpose of the data collection, any third parties to which the data may be transferred, the duration of the storage of the data, and the data subjects’ rights and how to use them. (Tietosuojavaltuutettu 2019.)

Thus, the study material consists of certain passages collected from six privacy statements translated from English into Finnish and from nine privacy statements originally written in Finnish. As the GDPR regulates the content of privacy statements, there are certain sections

that occur in each privacy statement regardless of the data controller. Thus, to ensure comparability, the following sections were collected from each privacy statement:

- (1) sections describing **the data collected and processed**
- (2) sections describing **the purpose of the data collection**
- (3) sections describing **the data subjects' rights**
- (4) possible **introduction**.

As privacy statements can vary in length quite significantly, the number of translated texts is lower than that of originally produced texts. However, the word count is roughly the same for both: 5,142 words in total for the translations and 4,707 words for the originally produced texts.

The study method of this thesis is comparative, with a quantitative focus. The purpose is to compare and contrast the translations and the originally produced texts with regard to the direct forms of address used. The comparison is carried out by utilising a slightly modified version of Eveliina Tolvanen's (2016) classification of direct forms of address. All direct forms of address were thus manually collected from the study material and classified into the following categories:

- (1) verbs in the second-person singular form
- (2) the personal pronoun '*sinä*' (English: 'you')
- (3) possessive suffixes
- (4) imperative verbs.

After collecting and classifying the forms of address into the categories above, certain calculations were made. The following figures were calculated for each individual text, on the one hand, and for the group of translations as a whole and the group of originally produced texts as a whole, on the other:

- (1) the total number of forms of address
- (2) the total number of forms of address under each class
- (3) the total percentage of forms of address in the whole text.

Additionally, it was calculated how many per cents of the total number of address forms each class represents in the group of translations and the group of originally produced texts. Finally,

to ensure comparability, the number of address forms per 1,000 words was calculated for both groups of texts. Based on the presented calculations, some conclusions could be drawn up, which are observed in the following section.

4 Analysis and Results

The analysis was conducted by calculating the figures described above in Section 3 and then comparing the figures between the translations and originally produced texts. Below is a table that shows the word count of the translations as compared to the originally produced texts, as well as the number and the percentage of address forms in both groups of texts:

	Translations	Originally produced texts
Word count in total	5,142	4,707
Number of address forms in total	594	148
Percentage of address forms in total	11.6	3.1

Table 1. Scope of the study material and the number and percentage of address forms.

As can be seen from the table, the translations and originally produced texts differ drastically when it comes to the frequency of forms of address in the material: 11.6% of the words in the translations are address forms, while the corresponding figure for the originally produced texts is 3.1%. Below is an authentic example of how the translations (T) and originally produced texts (O) are different in terms of addressing the reader:

- (1) (T) *Sinulla* on oikeus tarkastaa, mitä *sinua* koskevia tietoja rekisteriin on tallennettu. (K5)
(O) Henkilöllä on oikeus tarkastaa, mitä tietoja hänestä on talletettu asiakasrekisteriin. (R4)

In the example above, direct forms of address have been italicised. Both sentences express roughly the same thing, i.e. that the reader has the right to know what personal data is being stored. The translation uses a direct form of address, the personal pronoun ‘*sinä*’, in two different inflected forms. The originally produced text, however, uses an indirect, nominal form of address and addresses the reader as ‘*henkilö*’ (English: ‘person’).

In addition to the sheer volume of address forms, there are also differences in how the different classes of address forms are distributed in the study material. In order to gain a more detailed understanding, the next table will illustrate the volumes of each class in the material:

	Translations, address forms per 1,000 words	Originally produced texts, address forms per 1,000 words
Verb, 2 nd person singular	35	6
The personal pronoun ‘ <i>sinä</i> ’	18	7
Possessive suffix	60	17
Imperative verb	3	1
Total	116	31

Table 2. Number of address forms per 1,000 words in the material.

After considering the table above, it becomes quite apparent that forms of address appear more frequently in the translations. However, the most and least common classes of address forms appear to be the same in both groups of texts – possessive suffix is the most common and imperative verb is the least common of the classes. What seems somewhat unusual is the fact that the personal pronoun ‘*sinä*’ is the second most common class in the originally produced texts. As mentioned earlier, it is typical for Finnish to omit the personal pronoun and only use the verb inflection or possessive suffix to indicate address (see Section 2).

There is also some variation between individual texts. Somewhat surprisingly, the translations were quite similar when compared with each other. All texts follow the distribution of the different classes as described in Table 2 above, meaning that the most, second, third, and least common classes are the same in each text. Moreover, the percentage of address forms in each text is roughly 10%. The originally produced texts, on the other hand, contain quite a lot of variation. Some texts contain no direct address forms whatsoever, while others almost match the translations when looking at the percentage of direct address forms in the text. It is also noteworthy that the originally produced texts seem to prefer indirect forms of address over direct forms, but in the translations, indirect forms of address are virtually non-existent.

5 Concluding Remarks

The purpose of this study was to examine the differences and similarities in the use of direct address forms between privacy statements translated from English into Finnish and privacy statements originally written in Finnish. The hypothesis was that the translations would contain more direct forms of address as compared to the originally produced texts. Based on the analysis of the study material of six translations and nine originally produced texts, it is clear that direct forms of address are more common in the translations as compared to the originally produced texts. This could be a result of interference, i.e. source language impact on the language of the target text, as English uses direct forms of address more frequently than Finnish does. However, as the study was limited in nature and only focused on texts of a certain type, no broad generalisations can be made based on the results.

As this study focused solely on direct forms of address in the study material, it would be interesting to observe indirect forms of address in the same texts in the future. It would also be useful to conduct a similar study on a broader selection of texts, as the study material in this study was quite limited in size. All in all, the results of this study provide an indication of the way address forms are used in translations of privacy statements from English into Finnish as compared to privacy statements originally written in Finnish. The results also support earlier research on the general features of translated language, particularly Finnish.